

STIC-Biotech/ChemLib

79093

From: Portner, Ginny  
Sent: Wednesday, October 30, 2002 1:58 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/910,186  
Importance: High

Please search SEQ ID Nos 9 and 10, please back translate SEQ ID NO 10. Thanks.

Ginny Portner  
CM1, Art Unit 1645  
Room 7e13  
Mail box 7e12  
(703) 308-7543

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

RECEIVED  
OCT 30 2002  
STIC

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/4/02  
Date Completed: 11/8/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: 141  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 10/6  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into to two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

*The Pending database search results should not be left in the case because they contain data that is confidential.*



According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:

**Published\_Applications\_NA** contains nucleic acid sequences; the search results will have the extension **.rnpb**.

**Published\_Applications\_AA** contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

**Sequences in the PGPub database are public information; it is permissible to leave these results in the case.**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 13:45:26 ; Search time 55 Seconds  
(without alignments)  
922.504 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MIPPFNIFSYTNSLLKDI.....NYASLLESTSHWGFPVSE 450

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_101002.\*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
  - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2391	100.0	450	21	AAV77136
2	2391	100.0	450	22	AAV77136
3	2386	99.8	1291	20	AAV05814
4	2317	96.9	462	19	AAW68397
5	1599	66.5	399	21	AAV78982
6	867.5	35.3	451	22	AAV04093
7	801.5	33.5	451	19	AAW68398
8	589.5	24.7	449	21	AAV77137
9	589.5	24.7	449	22	AAV04094
10	586.5	24.5	837	21	AAV77140

11	586.5	24.5	1067	21	AAV93307	A manganese supero
12	586.5	24.5	1092	21	AAV93310	A manganese supero
13	586.5	24.5	1296	17	AAV93308	C. botulinum type
14	581	24.3	1070	21	AAV93308	A manganese supero
15	581	24.3	1095	21	AAV93311	A manganese supero
16	581	24.3	1291	19	AAW68392	Clostridium botuli
17	578.5	24.2	452	19	AAW68396	Clostridium botuli
18	577.5	24.2	1295	23	AAU99339	Clostridium botuli
19	576	24.1	451	19	AAW68395	Synthetic botulinu
20	575.5	24.1	449	21	AAV77139	Botulism toxin hea
21	575.5	24.1	449	22	AAV04167	Clostridium botuli
22	575.5	24.1	473	19	AAW68400	Botulism toxin hea
23	573	24.0	848	22	AAV04082	Botulism toxin hea
24	568.5	23.6	472	19	AAW68393	Clostridium botuli
25	565	23.6	472	19	AAW68393	A manganese supero
26	564	23.6	1059	21	AAV93309	A manganese supero
27	564	23.6	1084	21	AAV93312	C. botulinum C2 tr
28	558	23.3	1092	22	AAE07900	Native botulinum n
29	557.5	23.3	432	21	AAV77142	Botulism toxin hea
30	557.5	23.3	437	22	AAV04088	Type A neurotoxin
31	557.5	23.3	438	17	AAV95008	Clostridium botuli
32	557.5	23.3	438	19	AAW68389	Synthetic botulinu
33	557.5	23.3	438	21	AAV77134	Clostridium botuli
34	557.5	23.3	445	19	AAW68391	Type A neurotoxin
35	557.5	23.3	462	17	AAV95009	Clostridium botuli
36	557.5	23.3	462	19	AAW68390	Botulism toxin hea
37	556.5	23.3	434	22	AAV04089	Botulism toxin hea
38	556.5	23.3	435	22	AAV04090	Botulism toxin hea
39	553.5	23.1	419	22	AAV04095	C. botulinum C2 tr
40	551	23.0	1032	22	AAE07901	Clostridium botuli
41	550	23.0	472	19	AAW68394	Modified clostridi
42	548	22.9	685	22	AAV07893	Immunogenic type F
43	546	22.8	431	18	AAW09014	Botulism toxin hea
44	546	22.8	432	22	AAV04096	Botulism toxin hea
45	546	22.8	432	22	AAV04103	Botulism toxin hea

ALIGNMENTS

RESULT 1  
AAV77136  
ID AAV77136 standard; Protein: 450 AA.  
AC AAV77136;  
XX  
XX  
DT 08-MAY-2000 (first entry)  
DE Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.  
XX Botulinum neurotoxin; heavy chain; BoNT; serotype C;  
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
XX VEE; botulism; vaccine; diagnosis; drug screening.  
XX Clostridium botulinum.  
OS Synthetic.  
XX  
XX WO200002524-A2.  
XX  
XX 20-JAN-2000.  
XX  
XX 09-JUL-1999; 99WO-US15570.  
XX  
XX 10-JUL-1998; 98US-0092416.  
XX 12-MAY-1999; 99US-0133870.  
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
XX WPI: 2000-160827/14.  
XX N-PSDB; AA287214.  
XX

PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
PT toxin serotypes A-G, is used for inducing an immune response against  
PT botulinum .  
XX  
XX Claim 24; Page 41-42; 54pp; English.  
XX  
XX The invention relates to novel vaccines that induce a protective immune  
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
CC DNA construct comprising a vector, and at least one nucleic acid  
CC fragment comprising a C-terminal heavy chain fragment (HC) from BoNT  
CC serotypes A-G. In preferred embodiments of the invention, the vector is  
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
CC this vector results in the production of large amounts of a protein  
CC encoded by a sequence cloned into the replicon. The constructs are used  
CC to produce vaccines against botulinum. The proteins can also be used as  
CC diagnostic tools for the diagnosis of botulinum. The transformed host  
CC cells can be used to analyse the effectiveness of drugs and agents which  
CC inhibit toxin effects. The vaccine currently used against botulinum is  
CC dangerous and expensive to produce, and contains formalin, which is very  
CC painful for the recipient. Also, the vaccine is incomplete, in that only  
CC 5 of the 7 serotypes are represented in the formulation. The novel  
CC vaccine of overcomes these problems, as it is easily purified, and  
CC available in large quantities. It is also expressed in the lymph nodes  
CC for a better immune response. Sequences AAY7134-Y7139 represent  
CC synthetic BoNT Hc fragments used in the present invention. The DNA  
CC encoding these sequences had been optimised for codon usage for  
CC expression in yeast.  
XX  
XX Sequence 450 AA;

Query Match 100.0%; Score 2391; DB 21; Length 450;  
Best Local Similarity 100.0%; Pred. No. 9.6e-164;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTIPNIFSYNNLSKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
DB 1 MTIPNIFSYNNLSKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
QY 61 NPIFFDFKLSSGSDRGKVIQTONENIVNSMYESFSISFWIRINKWVSNLPGYIIDS 120  
DB 61 NPIFFDFKLSSGSDRGKVIQTONENIVNSMYESFSISFWIRINKWVSNLPGYIIDS 120  
QY 121 VKNNGWSGIGISNLFVTLKQNEDESEQINFSYDINNNAPGYNKWFVVTNNMGNNK 180  
DB 121 VKNNGWSGIGISNLFVTLKQNEDESEQINFSYDINNNAPGYNKWFVVTNNMGNNK 180  
QY 181 IYNGKLIDTIKVKELTGINFSTKIPINKIPDTGLITSDSDNNINWIRDFYIFAKELD 240  
DB 181 IYNGKLIDTIKVKELTGINFSTKIPINKIPDTGLITSDSDNNINWIRDFYIFAKELD 240  
QY 241 GKDINTLNSLOYTNVVDYWGNDLRYNKEYYVYNDYLNRYMYANSRQIVFNTRRNND 300  
DB 241 GKDINTLNSLOYTNVVDYWGNDLRYNKEYYVYNDYLNRYMYANSRQIVFNTRRNND 300  
QY 301 FNEGYKIIIRKIRGNNTDTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAI 360  
DB 301 FNEGYKIIIRKIRGNNTDTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAI 360  
QY 361 GLREQTKDINDNIIFQIOPMNTYVYASQIFKSNFNGENTSGICISGYFRLLGGDWYRH 420  
DB 361 GLREQTKDINDNIIFQIOPMNTYVYASQIFKSNFNGENTSGICISGYFRLLGGDWYRH 420  
QY 421 NYLVPVTKOGNYASLLESTSHWGFVPVSE 450  
DB 421 NYLVPVTKOGNYASLLESTSHWGFVPVSE 450

RESULT 2  
AAB04092  
ID AAB04092 standard; Protein; 450 AA.  
XX  
AC AAB04092;

XX 11-APR-2001 (first entry)  
XX Botulinum toxin heavy chain C-terminal sequence (serotype C).  
DE  
XX Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
KW infection.  
XX  
XX Synthetic.  
OS Clostridium botulinum.  
XX W0200067700-A2.  
XX 16-NOV-2000.  
XX 12-MAY-2000; 2000WO-US12890.  
XX 12-MAY-1999; 99US-0133865.  
PR 12-MAY-1999; 99US-0133866.  
PR 12-MAY-1999; 99US-0133867.  
PR 12-MAY-1999; 99US-0133868.  
PR 12-MAY-1999; 99US-0133869.  
PR 12-MAY-1999; 99US-0133873.  
PR 29-JUL-1999; 99US-0146192.  
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
FA  
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
PI  
XX WPI; 2001-016048/02.  
DR N-PSDB; AAA54486.  
XX  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulinum  
XX  
XX Claim 3; Fig 5b; 73pp; English.  
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulinum. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.

XX Sequence 450 AA;

Query Match 100.0%; Score 2391; DB 22; Length 450;  
Best Local Similarity 100.0%; Pred. No. 9.6e-164;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTIPNIFSYNNLSKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
DB 1 MTIPNIFSYNNLSKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
QY 61 NPIFFDFKLSSGSDRGKVIQTONENIVNSMYESFSISFWIRINKWVSNLPGYIIDS 120  
DB 61 NPIFFDFKLSSGSDRGKVIQTONENIVNSMYESFSISFWIRINKWVSNLPGYIIDS 120  
QY 121 VKNNGWSGIGISNLFVTLKQNEDESEQINFSYDINNNAPGYNKWFVVTNNMGNNK 180  
DB 121 VKNNGWSGIGISNLFVTLKQNEDESEQINFSYDINNNAPGYNKWFVVTNNMGNNK 180

QY 181 IYINGKLIDTIVKVELTGINFSTITEINKIPDTGLITSDSDNINMWIRDFVIFAKELD 240  
 DB 181 IYINGKLIDTIVKVELTGINFSTITEINKIPDTGLITSDSDNINMWIRDFVIFAKELD 240  
 QY 241 GKDINILFNSIQYTNVVDYWGNDLRYNKEYVMYNDYLNRYMYANSRQIVFVTRNNND 300  
 DB 241 GKDINILFNSIQYTNVVDYWGNDLRYNKEYVMYNDYLNRYMYANSRQIVFVTRNNND 300  
 QY 301 FNEGYKIIIRKIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360  
 DB 301 FNEGYKIIIRKIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360  
 QY 361 GLREOTKDINDNIIFOIPMNNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420  
 DB 361 GLREOTKDINDNIIFOIPMNNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420  
 QY 421 NYLVPTVKQGNVYASLLESTSTHWGFVPVSE 450  
 DB 421 NYLVPTVKQGNVYASLLESTSTHWGFVPVSE 450

RESULT 3  
 ID AAY05814  
 AC AAY05814; standard; Protein: 1291 AA.  
 XX AAY05814;  
 DT 02-AUG-1999 (first entry)  
 XX Non-toxic modified botulinum toxin rBONT/C.  
 DE Botulinum toxin; botulism; rBONT/C; vaccine; drug delivery;  
 KW mutant.  
 KW Clostridium botulinum.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH Misc-difference 229 /note= "His in native toxin"  
 FT Misc-difference 230 /note= "Glu in native toxin"  
 FT Misc-difference 233 /note= "His in native toxin"  
 FT Misc-difference 233 /note= "His in native toxin"  
 XX W0920306-A1.  
 XX 29-APR-1999.  
 XX 16-OCT-1998; 98WO-US21897.  
 XX 20-OCT-1997; 97US-0954302.  
 XX (UYJE-) UNIV JEFFERSON THOMAS.  
 XX Kiyatkin N, Maksymowych A, Simpson L;  
 XX WPI; 1999-302646/25.  
 DR N-PSDB; AAX25521.  
 XX Modified toxin useful for systemic delivery of oral vaccines and  
 PT therapeutic agents  
 XX Example 1; Page 31-33; 37pp; English.  
 XX The present sequence represents a modified serotype C botulinum  
 CC toxin, termed rBONT/C, in which amino acids His-229, Glu-230 and  
 CC His-233 of the native sequence are substituted by Gly, Thr and  
 CC Asn, respectively, i.e. the zinc binding motif (see AAY05817) of  
 CC the light chain holotoxin is modified, resulting in loss of  
 CC endoprotease activity. DNA coding for the modified botulinum toxin  
 CC (see AAX25521) was assembled from 3 separate toxin fragments using

CC PCR and site-directed mutagenesis. The modified recombinant  
 CC botulinum toxin maintains its ability to translocate from the gut  
 CC into the general circulation but is non-toxic. It can be used as  
 CC an oral vaccine for antigenic peptides including botulinum toxin  
 CC (i.e. an oral vaccine for botulism) or for the oral delivery of  
 CC other therapeutic agents to the general circulation.  
 XX Sequence 1291 AA;  
 SQ Query Match 99.8%; Score 2386; DB 20; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 8e-103;  
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TIPNIRSYTNNSLLKDIINEYFNININDSKILSLQNKNTLVDTSGYNAEVSSEGVDQLN 61  
 DB 843 TIPNIRSYTNNSLLKDIINEYFNININDSKILSLQNKNTLVDTSGYNAEVSSEGVDQLN 902  
 QY 62 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESTSISFWIRNKWVSNLPGYTIIDSV 121  
 DB 903 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESTSISFWIRNKWVSNLPGYTIIDSV 962  
 QY 122 KNSGWSIGIISNLFVTLKONEDSEQSINFSYDISNNAPYKNKWEFFVTNNMGMNMI 181  
 DB 963 KNSGWSIGIISNLFVTLKONEDSEQSINFSYDISNNAPYKNKWEFFVTNNMGMNMI 1022  
 QY 182 YINGKLIDTIVKVELTGINFSTITEINKIPDTGLITSDSDNINMWIRDFVIFAKELD 241  
 DB 1023 YINGKLIDTIVKVELTGINFSTITEINKIPDTGLITSDSDNINMWIRDFVIFAKELD 1082  
 QY 242 KDINILFNSLOYTNVVDYWGNDLRYNKEYVMYNDYLNRYMYANSRQIVFVTRNNND 301  
 DB 1083 KDINILFNSLOYTNVVDYWGNDLRYNKEYVMYNDYLNRYMYANSRQIVFVTRNNND 1142  
 QY 302 NEGKIIIRKIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
 DB 1143 NEGKIIIRKIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202  
 QY 362 LREOTKDINDNIIFOIPMNNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 421  
 DB 1203 LREOTKDINDNIIFOIPMNNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 1262  
 QY 422 YLVPTVKQGNVYASLLESTSTHWGFVPVSE 450  
 DB 1263 YLVPTVKQGNVYASLLESTSTHWGFVPVSE 1291

RESULT 4  
 AAW68397  
 ID AAW68397 standard; Protein: 462 AA.  
 XX AAW68397;  
 XX 07-DEC-1998 (first entry)  
 XX Clostridium botulinum type C1 toxin C fragment.  
 XX Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;  
 XX botulism; BotC.  
 XX Clostridium botulinum serotype C1 strain Stockholm.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH Peptide 1..21  
 FT /note= "N-terminal His tag"  
 XX W09808540-A1.  
 XX 05-MAR-1998.  
 XX 28-AUG-1997; 97WO-US15394.  
 XX 28-AUG-1996; 96US-0704159.  
 PR

```

XX PA (OPHI-) OPHIDIAN PHARM INC.
XX FI Thalley BS, Williams JA;
XX DR WPI; 1998-230234/20.
XX DR N-PSDB; AAV30588.
XX PT Host cell containing recombinant expression vector encoding
XX PT Clostridium botulinum type B or E toxin - useful to treat humans
XX PT and other animals at risk of intoxication with clostridial toxin
XX PS Example 45; Page 339-341; 428pp; English.
XX CC This is the amino acid sequence of the histidine-tagged C fragment
XX CC of Clostridium botulinum (Stockholm strain) type C1 neurotoxin,
XX CC encoded by a DNA sequence (see AAV30588) in plasmid pTH18b. This
XX CC vector is used to express Bacc soluble C fragment in Escherichia
XX CC coli host cells, and the recombinant C fragment was purified on an
XX CC affinity column. The invention relates to recombinant proteins
XX CC derived from C. botulinum toxins, especially type B and type E
XX CC toxins. Methods are provided which allow for the isolation of
XX CC soluble recombinant proteins free of significant endotoxin
XX CC contamination. Preferred hosts for production of recombinant
XX CC proteins are E. coli, insect cells and yeast cells. The
XX CC recombinant toxins are used as immunogens for the production of
XX CC vaccines and antitoxins that are useful in the treatment of humans
XX CC and animals at risk of intoxication with clostridial toxin.
XX PS Sequence 462 AA;
XX
XX Query Match 96.98; Score 2317; DB 19; Length 462;
XX Best Local Similarity 99.88; Pred. No. 2e-158;
XX Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 14 SLKDIINEYFNNDKILSLQNKNTLVDTSGYAEVSEEGDVLNPFPPDFKLGS 73
DB 26 ALLKDIINEYFNNDKILSLQNKNTLVDTSGYAEVSEEGDVLNPFPPDFKLGS 85
QY 74 GDRGKIVTQNTENIVNMYTESFSISFWIRINKWVSNLPGYTIIDSVKNSGWSIGIS 133
DB 86 GDRGKIVTQNTENIVNMYTESFSISFWIRINKWVSNLPGYTIIDSVKNSGWSIGIS 145
QY 134 NFLVFTLKQNEDESGINFSINFSIDSNAPGVNKFVTVTNMGNMKIYINGKLIDTIKV 193
DB 146 NFLVFTLKQNEDESGINFSINFSIDSNAPGVNKFVTVTNMGNMKIYINGKLIDTIKV 205
QY 194 KELTGINFSKTIITPEINKIPDTGLITSDSDINNMWIRDFYIFAKELGDKDINILFNSLQY 253
DB 206 KELTGINFSKTIITPEINKIPDTGLITSDSDINNMWIRDFYIFAKELGDKDINILFNSLQY 265
QY 254 TNVVDYWGNDLRYNKEYMYNIDYLNRYMYANSQIIVENTRRNNDNFEGYKIIIKRIR 313
DB 266 TNVVDYWGNDLRYNKEYMYNIDYLNRYMYANSQIIVENTRRNNDNFEGYKIIIKRIR 325
QY 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDYIAIGLREQTKDINDNI 373
DB 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDYIAIGLREQTKDINDNI 385
QY 374 IFQIQPMNNTYYASQIPKSNFNGENISGICSTGYFRGLGGDWYRNYLVPVYKQNYA 433
DB 386 IFQIQPMNNTYYASQIPKSNFNGENISGICSTGYFRGLGGDWYRNYLVPVYKQNYA 445
QY 434 SLESTSTHWGFPVYSE 450
DB 446 SLESTSTHWGFPVYSE 462
XX
RESULT 5
ID AAV78982
XX AAV78982 standard; Protein; 399 AA.
XX AC AAV78982;

```

```

XX 20-JUN-2000 (first entry)
XX DT
XX DE C. botulinum type D toxin amino acid sequence.
XX FI
XX KW Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
XX KW protect.
XX OS Clostridium botulinum.
XX PN W0200005252-A1.
XX PD 03-FEB-2000.
XX PF 20-JUL-1999; 99WO-IB01301.
XX PR 22-JUL-1998; 98ZA-0006538.
XX FA (AGRI-) AGRIC RES COUNCIL.
XX PI De Bruyn EE, Botha AD;
XX DR WPI; 2000-205375/18.
XX DR N-PSDB; AA298630, AA298631.
XX PT Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin
XX PT type D, useful in vaccines for protection against botulism, comprises
XX PT at least one amino acid mutation not present in the wild type D
XX PT neurotoxins -
XX PS Claim 3, 4; Page 54-57; 66pp; English.
XX
XX This sequence represents the amino acid sequence of a synthetic
XX non-toxic immunogenic derivative of Clostridium botulinum type D toxin
XX (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle and
XX sheep, and usually results in the death of the affected or poisoned
XX animal. The non-toxic immunogenic fragments of the C. botulinum
XX neurotoxin are useful in vaccines to protect animals (e.g. humans,
XX cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic
XX fragments can be produced relatively simply and inexpensively
XX (specifically by fermentation techniques). As the fragments are not
XX toxic the risk to production staff is reduced.
XX
XX Sequence 399 AA;
XX
XX Query Match 66.5%; Score 1589; DB 21; Length 399;
XX Best Local Similarity 76.08; Pred. No. 3.3e-106;
XX Matches 307; Conservative 33; Mismatches 56; Indels 8; Gaps 4;
XX
QY 50 AEVSEGDVQLNPIFPDFKLGSSEDRGKVIVTQNTENIVNMYTESFSISFWIRINKWV 109
DB 1 AEVRVGVQLNPIFPDFKLGSSEDRGKVIVTQNTENIVNMYTESFSISFWIRINKWV 60
QY 110 SNLPGYTIIDSVKNSGWSIGISNLFVTLKQNEDESGINFSIDSNAPGVNKFV 169
DB 61 SNLPGYTIIDSVKNSGWSIGISNLFVTLKQNEDESGINFSIDSNAPGVNKFV 120
QY 170 TVTNMGNMKIYINGKLIDTIKVYKELTGINFSKTIITPEINKIPDTGLITSDSDINMWI 229
DB 121 ITTNMGNMKIYINGKLIDTIKVYKELTGINFSKTIITPEINKIPDTGLITSDSDINMWI 180
QY 230 RDFYIFAKELGDKDINILFNSLQYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYANSRQ 289
DB 181 RDFYIFAKELGDKDINILFNSLQYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYANSRQ 240
QY 290 IVFNTRRNNDNFEGYKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYA 349
DB 241 IVFNTRRNNDNFEGYKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYA 295
QY 350 DNHST-EDIYAGLREQTKD-INDNIIFQIQPMNNTYYASQIPKSNFNGENISGICSIG 407
DB 296 PSRNLGTDLVPLGALDQPMDEIRKYGSIIFQPCNTFDYASQLFLSSNATNRLGILSIG 355

```



.....

OV 350 DNHSTEDIYAIGLREQTKDIND--NIIFQIQPMNNTYYASQIFKSNFNGENISGICSIG 40

Gaps 17;

```

Db 362 DTATTN-----KERTIKISSGNRFNQVVMNSVGNCTMNFKN-NGNN-----IG 407
QY 408 TYRFR---LGGDWYRHNLVPTVKQGNVASLLESTH---WGTV 446
Db 408 LLGFKADTVVASTWY-----YTHMRDHTNSGCFWNFI 440

RESULT 9
AAB04094
ID AAB04094 standard; Protein; 449 AA.
XX AAB04094;
AC AAB04094;
XX
XX 11-APR-2001 (first entry)
DE Botulinism toxin heavy chain C-terminal sequence (serotype E).
XX
XX Botulinism toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
XX Synthetic.
OS Clostridium botulinum.
XX
XX WO2000067700-A2.
PN
XX
XX 16-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US12890.
XX
XX 12-MAY-1999; 99US-0133865.
XX
XX 12-MAY-1999; 99US-0133866.
XX
XX 12-MAY-1999; 99US-0133867.
XX
XX 12-MAY-1999; 99US-0133868.
XX
XX 12-MAY-1999; 99US-0133869.
XX
XX 12-MAY-1999; 99US-0133870.
XX
XX 29-JUL-1999; 99US-0146192.
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
PI
XX WPI; 2001-016048/02.
XX
XX N-PSDB; AAA54488.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
XX Claim 3; Fig 7b; 73pp; English.
XX
XX Botulinism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from clostridium botulinum serotypes A-G.
XX
XX Sequence 449 AA;
SQ
Query Match 24.7%; Score 589.5; DB 22; Length 449;
Best Local Similarity 32.6%; Pred. No. 2.3e-34;

```

```

Matches 152; Conservative 90; Mismatches 159; Indels 65; Gaps 17;
QY 2 TIPFNIFYSNLLKDIINEYFNNTINDSKILQNRKNTLVDTSYNAEVSSEGDVQLN 61
Db 19 SIPEKLSYDDDKILISYFNKFKRIKSSSVLMRYKNDKYDTSYDSDNININGDVYKY 78
QY 62 PIFPDPKLGSSGDRGVIVTONENIVNMYESFISFWIRI---NKWYSNLPGVYI 117
Db 79 PTNKNQF--GIYNDKLTSLNDSYDIIDNKYKNSISFWVRIPNYDNKIYNNVEYTI 136
QY 118 IDSVK--NNSGWSIGIISNLFVFTLKQNEDESEQINSFYDISNAPGY--NKWFFVTVNNM 175
Db 137 INCMRDNSGWKYSNLNHNIEIWTIQDNAGINQKLAFTNYGNANGISDYINKWIFVITNDR 196
QY 176 MGNMKIYINGKLIDTIKVKELTGINFSKTITPEINKIPDTGLITSDSDNINMWIRDFYIF 235
Db 197 LGDSKLYINGNLIDQKSLNLGHNHSDNLFKIVNCSYTRI-----GIRYFNIF 247
QY 236 AKELGDKDINILFNSLOYTNVVDYMGNDLRYNKYMYMVDYLNRYM--YANSQIIVFN 293
Db 248 DKELDEIQTLYSNPEPTNLIKDFWGNVLLYDKVYLLNVLKPNFIDRRKXDSLISINN 307
QY 294 TRRN---NNDENEGYKIIKRI--RGNTNDRVRGGDILYEDTINNKAYNLPMKNETMYA 349
Db 308 IRSTILLANRLYSGIKVQIQRVANSSTNDNLVRKNDQVYINFVASKT--HLF---PLTA 361
QY 350 DNHSTEDIYAIGLREQTKDIND--NIIFIQPMNNTYVYASQIFKSNFNGENISGICISG 407
Db 362 DTATTN-----KEKTIKISSGNRFNQVVMNSVGNCTMNFKN-NGNN-----IG 407
QY 408 TYRFR---LGGDWYRHNLVPTVKQGNVASLLESTH---WGTV 446
Db 408 LLGFKADTVVASTWY-----YTHMRDHTNSGCFWNFI 440

RESULT 10
AAY77140
ID AAY77140 standard; Protein; 837 AA.
XX AAY77140;
XX
XX 08-MAY-2000 (first entry)
DT
XX Native botulinum neurotoxin serotype A (BoNTA).
DE
XX Botulinum neurotoxin; heavy chain; BoNT; serotype A;
KW Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening.
XX
XX Clostridium botulinum.
XX
XX Key Location/Qualifiers
XX Misc-difference 837
XX /note= "Apparently encoded by GGAITGGGAG AAGGCCACT G"
XX
XX WO200002524-A2.
XX
XX 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15570.
XX
XX 10-JUL-1998; 98US-0092416.
XX
XX 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
PI
XX WPI; 2000-160827/14.
XX
XX N-PSDB; AA287218.
XX
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against

```



PT botulinum -  
PS Example 3; Page 49; 54pp; English.  
XX  
XX The invention relates to novel vaccines that induce a protective immune  
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
CC DNA construct comprising a vector and at least one nucleic acid  
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT  
CC serotypes A-G. In preferred embodiments of the invention, the vector is a  
CC Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this  
CC vector results in the production of large amounts of a protein encoded by  
CC a sequence cloned into the replicon. The constructs are used to produce  
CC vaccines against botulinum. The proteins can also be used as diagnostic  
CC tools for the diagnosis of botulinum. The transformed host cells can be  
CC used to analyse the effectiveness of drugs and agents which inhibit toxin  
CC effects. The vaccine currently used against botulinum is dangerous  
CC and expensive to produce, and contains formalin, which is very painful  
CC for the recipient. Also, the vaccine is incomplete, in that only 5 of  
CC the 7 serotypes are represented in the formulation. The novel vaccine  
CC of overcomes these problems, as it is easily purified, and available in  
CC large quantities. It is also expressed in the lymph nodes for a better  
CC immune response. The present sequence represents the native BoNTA heavy  
CC chain used in an exemplification of the present invention.  
XX  
SQ Sequence 837 AA;  
Query Match 24.5%; Score 586.5; DB 21; Length 837;  
Best Local Similarity 31.6%; Pred. No. 8.2e-34;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
QY 3 IPNFIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGDVQLNP 62  
DB 396 IPFQLSKYVDNQRLSTFTTEYIKNIINTSLNRYESNHLIDLSRYASKINIGSKVNFDP 455  
QY 63 IPPFD---FKLGSSEGRGVIVTONENIYVNSMYESFSFSFWIRNKVSNLP---GYT 116  
DB 456 IDKQIQIOLFNLESS-----KIEVLKNAIYNSMYENFSTFSFWIRPKYFNSISLNNEYT 510  
QY 117 IDSVKNNSGWSIGIISNLFVTLKQNEDESEQISNFSYDISNAPGY-NKWEFFVYTNMK 175  
DB 511 IINCENNSGKWSVSLNGEIIWTLQDTQEIQRVVKYSQMINISDYINRWIFVTITNR 570  
QY 176 MGNKIYINGKLIDTIKVKELTGFNFSTFTFEINKNTIPDGLITSDSDINMWIRDFYIF 235  
DB 571 LNSKIYINGRLIDQKPTISLNGTSHASNNIMFKLDGCRDT-----HRYIWKYFNLF 622  
QY 236 AKELDGKDNILFNSLOQYTNVVDYWGNDLRYNKKEYYMNIDYLNRYMYANS----- 287  
DB 623 DKELNEKEIKLDYDNQNSGILKDFWDYLDQYDKPYMLNLYDPNKYVDVNNVYRGYMY 682  
QY 288 ----RQIVFNTNR-RNNNDFNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFM 342  
DB 683 LKGPGRGVTMTNIVNLSLYRGTFEIKKYASGNKDNIVRNDRVYINVVVKNKEYRL-- 740  
QY 343 KNETMYADNHSSTEDYATGLREQTKDINDIIFIQIPMNNTYVYASQIFKSNPENGINS 402  
DB 741 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 793  
QY 403 ICSITVFRIGGDYRNYNLVPTVVKOGNTASLLEST-----THWGFVPVSE 450  
DB 794 IGFIFGHOF-----NNIAKLVASWNYNRQIERSRSLGCSWEFIPVDD 836  
RESULT 11  
AA93307  
ID AA93307 standard; protein; 1067 AA.  
XX  
AC AA93307;  
XX  
DT 04-SEP-2000 (first entry)  
XX  
DE A manganese superoxide dismutase (Mn-SOD) construct.

XX Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
KW neuronal cell targeting component; NCTC; neuronal disease;  
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
KW Huntington's disease; motor neurone disease;  
KW botulinum neurotoxin serotype A.  
XX  
OS Synthetic.  
OS Bacillus stearothermophilus.  
OS Clostridium botulinum.  
XX  
PN WO200028041-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-GB03699.  
XX  
PR 05-NOV-1998; 98GB-0024282.  
XX  
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
PI Shone CC, Sutton JM, Hallis B, Silman N;  
XX WPI; 2000-376553/32.  
XX  
XX Novel composition, comprising superoxide dismutase linked by a  
PT cleavable linker to a neuronal cell targeting component useful for  
PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
PS Disclosure; Page 41-44; 65pp; English.  
XX  
CC The present sequence represents a construct of the invention, comprising  
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that  
CC can be cleaved by thrombin, and a heavy chain derived from botulinum  
CC neurotoxin serotype A. The specification describes a composition for  
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,  
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).  
CC This component has a domain that binds to a neuronal cell and a  
CC domain that translocates the SOD of the composition into the neuronal  
CC cell. After translocation, the linker is cleaved to release the SOD.  
CC The composition is useful for treating neuronal diseases caused or  
CC augmented by oxidative stress, such as ischemic stroke, trauma,  
CC Parkinson's disease, Huntington's disease and motor neurone diseases.  
XX  
SQ Sequence 1067 AA;  
Query Match 24.5%; Score 586.5; DB 21; Length 1067;  
Best Local Similarity 31.8%; Pred. No. 1.1e-33;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
QY 3 IPNFIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGDVQLNP 62  
DB 620 IPFQLSKYVDNQRLSTFTTEYIKNIINTSLNRYESNHLIDLSRYASKINIGSKVNFDP 679  
QY 63 IPPFD---FKLGSSEGRGVIVTONENIYVNSMYESFSFSFWIRNKVSNLP---GYT 116  
DB 680 IDKQIQIOLFNLESS-----KIEVLKNAIYNSMYENFSTFSFWIRPKYFNSISLNNEYT 734  
QY 117 IDSVKNNSGWSIGIISNLFVTLKQNEDESEQISNFSYDISNAPGY-NKWEFFVYTNMK 175  
DB 735 IINCENNSGKWSVSLNGEIIWTLQDTQEIQRVVKYSQMINISDYINRWIFVTITNR 794  
QY 176 MGNKIYINGKLIDTIKVKELTGFNFSTFTFEINKNTIPDGLITSDSDINMWIRDFYIF 235  
DB 795 LNSKIYINGRLIDQKPTISLNGTSHASNNIMFKLDGCRDT-----HRYIWKYFNLF 846  
QY 236 AKELDGKDNILFNSLOQYTNVVDYWGNDLRYNKKEYYMNIDYLNRYMYANS----- 287  
DB 847 DKELNEKEIKLDYDNQNSGILKDFWDYLDQYDKPYMLNLYDPNKYVDVNNVYRGYMY 906  
QY 288 ----RQIVFNTNR-RNNNDFNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFM 342  
DB 907 LKGPGRGVTMTNIVNLSLYRGTFEIKKYASGNKDNIVRNDRVYINVVVKNKEYRL-- 964

Qy 343 KNETMYADNHSTEDIYAIGUREQTKDINDNIIFQIQPMNTYYIASQIFKSNFNGENISG 402  
 Db 965 -----ATNASQAGVERKILSALETPDVGNSLQVVMKSKNDQGITKNC-KMNLQDNNGND 1017  
 Qy 403 ICSIGTYRFRLLGGDWYRHNLYPTVYKQGNVSLLESTSS---THWGFVPVSE 450  
 Db 1018 IGFIFGHQF-----NNIAKLVSANWTYNROIERSRSLTGLCSWEFIPVDD 1060

RESULT 12  
 AAY93310  
 ID AAY93310 standard; protein; 1092 AA.  
 AC AAY93310;  
 XX  
 DT 04-SEP-2000 (first entry)  
 Db  
 DE A manganese superoxide dismutase (Mn-SOD) construct.  
 XX  
 KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
 KW neuronal cell targeting component; NCIC; neuronal disease;  
 KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
 KW Huntington's disease; motor neurone disease;  
 KW botulinum neurotoxin serotype A.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 OS Bacillus stearothermophilus.  
 OS Clostridium botulinum.  
 XX  
 PN WO200028041-A1.  
 PD 18-MAY-2000.  
 XX  
 PF 05-NOV-1999; 99WO-GB03699.  
 XX  
 PR 05-NOV-1998; 98GB-0024282.  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 PI Shone CC, Sutton JM, Hallis B, Silman N;  
 XX WPI; 2000-376553/32.  
 DR  
 PT Novel composition, comprising superoxide dismutase linked by a  
 PT cleavable linker to a neuronal cell targeting component useful for  
 PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
 PS Disclosure; Page 51-54; 65pp; English.  
 XX  
 CC The present sequence represents a construct of the invention, comprising  
 CC a mitochondrial leader sequence from human manganese superoxide  
 CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker  
 CC that can be cleaved by thrombin, and a heavy chain derived from  
 CC botulinum neurotoxin serotype A. The specification describes a  
 CC composition for delivery of SOD to neuronal cells. The composition  
 CC comprises SOD linked, by a cleavable linker, to a neuronal cell  
 CC targeting component (NCIC). This component has a domain that binds  
 CC to a neuronal cell and a domain that translocates the SOD of the  
 CC composition into the neuronal cell. After translocation, the linker  
 CC is cleaved to release the SOD. The composition is useful for treating  
 CC neuronal diseases caused or augmented by oxidative stress, such as  
 CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and  
 CC motor neurone diseases.  
 XX  
 SQ Sequence 1092 AA;  
 Query Match 24.5%; Score 586.5; DB 21; Length 1092;  
 Best Local Similarity 31.6%; Pred. No. 1 ie-33;  
 Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSVTNNLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSESGDVQLNP 62

Db 645 IPFOLSKYVDNQRLSTFTTEYIKNIINTSILNRYESNHLIDLSYASKINIGSKVNDP 704  
 Qy 63 IFPPD---FKLSSGEDRGKVIYVTQENIYVNSMYESFISFWIRINKVWSNLP---GYT 116  
 Db 705 IDKNOIQLEFNLESS---KIEVLKNAIVNSMYENFSTFWIRIPKYPFNSISLNEYT 759  
 Qy 117 IIDSVKNSGWSIGIISNLFVFLTKQNEDESEQSINFSYDISNAPGY-NKWEFFVTNNM 175  
 Db 760 IINMENNNGKVSLSNGEIIWLTQTEIKQKVYKYSQMINISDYINRWIFVITNNR 819  
 Qy 176 MGNMKIYINGKLIDITIKVKELTGINFSKTIITFEINKIPDTGLITSDSDINMWIRDFYIF 235  
 Db 820 LNSKIYINGRLIDQKPIPSNLGNHASNNIMFKLDCRDT-----HRYIWKYFNL 871  
 Qy 236 AKELDGKDINILFNSLOYTNVVDYWGNDLRYNKKEYMYNIDYLNRYMYANS----- 287  
 Db 872 DKELNEKEIKDLYDNQNSGILKDFWGDYQYDKPYMLNLYDPNKYVDVNNVGIRGYM 931  
 Qy 288 ---RQIVFNTR-RNNNDFNEGKIIKIRGNTNTRVRGGDILYFDMTINKKAYNLFM 342  
 Db 932 LKGRGSGVMNTIYLSLSLYRGKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRL-- 989  
 Qy 343 KNETMYADNHSTEDIYAIGUREQTKDINDNIIFQIQPMNTYYIASQIFKSNFNGENISG 402  
 Db 990 -----ATNASQAGVERKILSALETPDVGNSLQVVMKSKNDQGITKNC-KMNLQDNNGND 1042  
 Qy 403 ICSIGTYRFRLLGGDWYRHNLYPTVYKQGNVSLLESTSS---THWGFVPVSE 450  
 Db 1043 IGFIFGHQF-----NNIAKLVSANWTYNROIERSRSLTGLCSWEFIPVDD 1085

RESULT 13  
 AAR95010  
 ID AAR95010 standard; Protein; 1296 AA.  
 AC AAR95010;  
 XX  
 DT 09-JUL-1996 (first entry)  
 DE C. botulinum type A neurotoxin.  
 XX  
 KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.  
 XX Clostridium botulinum.  
 OS  
 PN WO9612802-A1.  
 XX  
 PD 02-MAY-1996.  
 XX  
 PF 23-OCT-1995; 95WO-US13737.  
 XX  
 PR 07-JUN-1995; 95US-0480604.  
 PR 24-OCT-1994; 94US-0329154.  
 PR 16-MAR-1995; 95US-0405496.  
 PR 14-APR-1995; 95US-0422711.  
 XX  
 PA (OPHI-) OPHIDIAN PHARM INC.  
 XX  
 PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
 PI Williams JA;  
 XX  
 DR WPI; 1996-230603/23.  
 DR N-PSDB; AAT29244.  
 XX  
 PT Fusion proteins comprising non-toxin protein and part of toxin -  
 PT useful to form anti-toxins against Clostridium botulinum type A, and  
 PT C. difficile type toxins, and to treat C. difficile intoxication,  
 PT partic. diarrhoea  
 XX  
 PS Claim 4; Page 344-350; 434pp; English.  
 XX Clostridium botulinum type A neurotoxin (AAR95010) is processed to form

CC a dimer composed of a light and a heavy chain. It is the product of  
 CC the type A neurotoxin gene (AAR29244). The 50 kDa C-terminal portion  
 CC of the heavy chain, or C fragment (see also AAR95008), was produced  
 CC using a synthetic gene (AAR29245) having codon usage altered to improve  
 CC expression in *Escherichia coli*. Fusion proteins of the type A toxin  
 CC or C fragment, with e.g. maltose binding protein or polyhistidine  
 CC affinity tag (see also AAR95008), are used to generate neutralising  
 CC antitoxins and in vaccine compns.

XX Sequence 1296 AA;

Query Match 24.5%; Score 586.5; DB 17; Length 1296;  
 Best Local Similarity 31.6%; Pred. No. 1.4e-33;  
 Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IFFNIFSTNNLLKDIINFFNNINDSKILSQNRKNTLVDTSGYNVSEEGDVQLNP 62  
 Db 849 IFFOLSKYVNDQRLSTFTTYIKNIINTSLNRYNSHLIDUSRYASIKNGSKVNDP 908  
 Qy 63 IFFPD---FKLGGSGEDRGKVIYVTONENIYNGWYSEFSTFWIRINKVSNLP---GYT 116  
 Db 909 IDKQIQIQLFNLESS-----KIEVLKNAIYNSWYENFSTFWIRIPKVFNSISLNNEYT 963  
 Qy 117 IDSVKNNSGISGIIISNLFVLKQNEDESEQINFSYDISNNAPGY-NKWFVVTNNM 175  
 Db 964 IINCENNSGKVSGLNYGEIITLQDTQEIQRVVKYQMINISDIYINRWIFVTIINR 1023  
 Qy 176 MGNKIYINGKLIDITIKVKELTGINFESKTIITFEINKIPDTGLITSDSDNINMWIRDFYF 235  
 Db 1024 LNNKIYINGKLIDITIKVKELTGINFESKTIITFEINKIPDTGLITSDSDNINMWIRDFYF 235  
 Qy 236 AKELDGKIDINILFNSLQYTNVVDYNGNDRYNKEYYMWIDVLYNMYVANS----- 287  
 Db 1076 DRELKEKEIKLDYDQNSGLKDFWGDYLDYQDKPYTLMNLYDPNKYVDYNNVYIRGYM 1135  
 Qy 288 ---RQIVENTR-RNNDFNRYGKIILIKRIGNTDPRVRGGDILYFDMTINKAYNLFM 342  
 Db 1136 LKGRGQVWNTIYNSLSYKGTFLIKKASGNKDNIVRNDVYINVVYKKEYL-- 1193  
 Qy 343 KNETMYADNHSTEDIYAGLEQTKDINDNIIFQIOPMNTYYVYASQIFKSNFNGENISG 402  
 Db 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVYVVKSKNDQGITNRC-KMNLQDNGND 1246  
 Qy 403 ICSIGTYFRLLGGWYRHNLYVPTVKQGNVSLSTST---TWGFPVPS 450  
 Db 1247 IGFIFGHQF-----NNIAKLVASWYNRQIERSRILGCSWEEIPVDD 1289

## RESULT 14

AA93308

ID AAY93308 standard; protein; 1070 AA.

XX

AC AAY93308;

XX

DT 04-SEP-2000 (first entry)

XX

DE A manganese superoxide dismutase (Mn-SOD) construct.

XX

KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;

KW neuronal cell targeting component; NCTC; neuronal disease;

KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;

KW Huntington's disease; motor neuron disease;

KW botulinum neurotoxin serotype B.

XX

OS Synthetic.

OS *Bacillus stearothermophilus*.

OS *Clostridium botulinum*.

XX

PN WO200028041-A1.

XX

PD 18-MAY-2000.

XX

PF 05-NOV-1999; 99WO-GB03699.

05-NOV-1998; 98GB-0024282.

(MICR-) MICROBIOLOGICAL RES AUTHORITY.

Shone CC, Sutton JM, Hallis B, Silman N;

WPI; 2000-376553/32.

Novel composition, comprising superoxide dismutase linked by a  
 cleavable linker to a neuronal cell targeting component useful for  
 delivering superoxide dismutase to neuronal cells to treat ischemia -  
 Disclosure; Page 45-47; 65pp; English.

The present sequence represents a construct of the invention, comprising  
 a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that  
 can be cleaved by thrombin, and a heavy chain derived from botulinum  
 neurotoxin serotype B. The specification describes a composition for  
 delivery of SOD to neuronal cells. The composition comprises SOD linked,  
 by a cleavable linker, to a neuronal cell targeting component (NCTC).  
 This component has a domain that binds to a neuronal cell and a  
 domain that translocates the SOD of the composition into the neuronal  
 cell. After translocation, the linker is cleaved to release the SOD.  
 The composition is useful for treating neuronal diseases caused or  
 augmented by oxidative stress, such as ischemic stroke, trauma,  
 Parkinson's disease, Huntington's disease and motor neuron diseases.

Sequence 1070 AA;

Query Match 24.3%; Score 581; DB 21; Length 1070;  
 Best Local Similarity 32.9%; Pred. No. 2.8e-33;  
 Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;

Qy 3 IFFNIFSTNNLLKDIINFFNNINDSKILSQNRKNTLVDTSGYNVSEEGDVQLNP 62  
 Db 615 MPFDLSITVNTDITLIEMFNKYNSEILNLIILRYKDNLDLSGYGAKVYVDGVELND 674

Qy 63 IFFPDKLGGSGEDRGKVIYVTONENIYNSWYSEFSTFWIRINKW---VSNL--PGYT 116

Db 575 --KNQFKLTSSA--NSKIRVTONIIFNSVFLDFSVFWIRIPKYNKNGIQNYIHNEYT 730

Qy 117 IDSVKNNSGISGIIISNLFVTLKQNEDESEQINFSYDISNNAPGY-NKWFVVTNNM 175

Db 731 IINCENNSGKVSGLNYGEIITLQDTQEIQRVVKYQMINISDIYINRWIFVTIINR 789

Qy 176 MGNKIYINGKLIDITIKVKELTGINFESKTIITFEINKIPDTGLITSDSDNIN-MWIRDFYI 234

Db 790 LNNKIYINGKLESNTDIKIDIREVIAEIIKLD-----GDIDRTQIFWYFST 840

Qy 235 FAKELDGKIDINILFNSLQYTNVVDYNGNDRYNKEYYMWIDVLYNMYVANS-----RQI 290

Db 841 ENTELSQSNIEERYKIQSYELKDFWGNPLMYNKEYYMFNAGNKNSYIKLKSDSPVGEI 900

Qy 291 VENTRRNN-----DENEGYKIIIRIGNT---NDTRVRGGDILYFD-MTINK-- 336

Db 901 LTRSKYNQNSKIYINRDLYIGKEFIIR-RKSNQSOSINDIVRVEDIYILDFNNGEVRV 959

Qy 337 -AYNLFEMKNETMYADNHSTEDIYAGLEQTKDINDNIIFQIOPMNTYYVYASQ-IFKSN 394

Db 960 YTYKFKKEE-----EKLFLAPISDSDEFYN---TIQREYDEQPTSCQLLKK- 1006

Qy 395 FNGENISGICSIGYRFRLLGG-----DWRHNYLVPTVKQGNVSLSTST 441

Db 1007 -DEESTDEIGLIGHRFYESGIVFEYKDYFCISKWY-----LKEVKKRPYKLNK---GC 1057

Qy 442 HWGFVPVSE 450

Db 1058 NWQFIPKDE 1066

RESULT 15

AA93311



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:43:01 ; Search time 26 Seconds  
(without alignments)  
509.243 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 WTIPNIFSTNNSLLKDI.....NYASLLESTSHWGFPVSE 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	586.5	24.5	1296	1	US-08-480-604A-28
2	586.5	24.5	1296	2	US-08-405-496A-28
3	586.5	24.5	1296	4	US-08-915-136-28
4	557.5	23.3	438	1	US-08-480-604A-23
5	557.5	23.3	438	2	US-08-405-496A-23
6	557.5	23.3	438	4	US-08-915-136-23
7	557.5	23.3	462	1	US-08-480-604A-26
8	557.5	23.3	462	2	US-08-405-496A-26
9	557.5	23.3	462	4	US-08-915-136-26
10	539.5	22.6	1169	4	US-08-235-829-20
11	425	17.8	853	4	US-08-913-880C-17
12	425	17.8	858	4	US-08-913-880C-16
13	425	17.8	860	4	US-08-913-880C-15
14	425	17.8	862	4	US-08-913-880C-14
15	425	17.8	865	4	US-08-913-880C-13
16	425	17.8	866	4	US-08-913-880C-12
17	425	17.8	874	4	US-08-913-880C-11
18	425	17.8	875	4	US-08-913-880C-10
19	425	17.8	1315	1	US-08-913-880C-1
20	423.5	17.7	618	1	US-08-668-381A-5
21	420	17.6	452	1	US-07-618-312A-2
22	420	17.6	452	1	US-07-618-312A-4
23	420	17.6	452	1	US-08-280-228-2
24	420	17.6	452	1	US-08-280-228-4
25	397	16.6	452	1	US-08-110-786A-8
26	148.5	6.2	977	4	US-09-206-942-53
27	148.5	6.2	983	4	US-09-206-942-51

28	143.5	6.0	599	2	US-08-910-551B-2	Sequence 2, Appli
29	142.5	6.0	476	4	US-09-316-083-3	Sequence 3, Appli
30	142.5	6.0	3135	1	US-08-323-170B-2	Sequence 2, Appli
31	142.5	6.0	3135	4	US-08-954-441-2	Sequence 2, Appli
32	138.5	5.8	2366	1	US-08-480-604A-10	Sequence 10, Appl
33	138.5	5.8	2366	2	US-08-405-496A-10	Sequence 10, Appl
34	138.5	5.8	2366	4	US-08-915-136-10	Sequence 10, Appl
35	138.5	5.8	2366	4	US-08-957-310-10	Sequence 10, Appl
36	128	5.4	1430	3	US-09-008-172-2	Sequence 2, Appli
37	128	5.4	1430	4	US-09-210-361-6	Sequence 6, Appli
38	128	5.4	1430	4	US-09-740-274-6	Sequence 6, Appli
39	127.5	5.3	1529	2	US-08-728-470-10	Sequence 10, Appl
40	127.5	5.3	1529	4	US-08-719-641-10	Sequence 10, Appl
41	127.5	5.3	1545	4	US-08-296-791-4	Sequence 4, Appli
42	127.5	5.3	1545	5	PCT-US95-10661A-4	Sequence 4, Appli
43	127	5.3	1702	4	US-08-296-791-5	Sequence 5, Appli
44	127	5.3	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
45	126	5.3	2710	1	US-08-480-604A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-480-604A-28  
; Sequence 28, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027

```
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-28

Query Match          24.5%; Score 586.5; DB 1; Length 1296;
Best Local Similarity 31.6%; Pred. No. 2.6e-39;
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPNFIYTNNSLKDIINFYNNINDSKILSONKKNLVDTSYGNAEVSEGDVQLNP 62
Db 849 IPFOLSKYVDNQRLSTFTYEIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNDP 908
Qy 63 IFPPD---FKLGSSGDRGKVIYVTONENIVYNSMYESFSISFWIRINKVSNLP---GYT 116
Db 909 IDKQIQIOLFNESS-----KIEVLKNAIVYNSMYENFSTFWIRIPKYNFISLNNEYT 963
Qy 117 IDSVKNNSGWSIGIISNLFVFTLKQNESEQSINFSYDISNNAPGY-NKWFVTVTNMM 175
Db 964 IINCMMNSGKVSILNLTQDTQEIQRVWFYKYSQMINISDIYNRWIFVTITNR 1023
Qy 176 MGNKVIYNGKLIDTIKVKELTGINSKTTTFEINKIPDTGLTSDNINMWIRDFYIF 235
Db 1024 LNNKIYINGRLIDQKPIISNLGNHASNNIMFKDGCDDT-----HRYIWKYFNLF 1075
Qy 236 AKELDGKDINILFNSLOYTNVVDYNGNDLRYNKYMYNIDYLNRYMYANS----- 287
Db 1076 DKELNEKEIKDLYDNQNSGILKDFWGDYLOYDKPYMNLNLYDPNKYVDVNVGIRGYM 1135
Qy 288 ----RQIVFNTR-RNNDFNEGKIIIRKIRGTNTRVRGGDILYFDTINNKAYNLFM 342
Db 1136 LKGRPGSVMTNIVYNSLSYRGTRFIKKYASGNKNIVNNDRVINVVVNKEVRL-- 1193
Qy 343 KNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISG 402
Db 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 1246
Qy 403 ICSTGTFRFLGGDWYRHNLVPTVKOGNTASLLEST-----THWGFVPVSE 450
Db 1247 IGFIFGHQF-----NNIAKLVASNWNRYNQIERSSTRTLCGSWEFIPVDD 1289

RESULT 3
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
```

```
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-28

Query Match          24.5%; Score 586.5; DB 1; Length 1296;
Best Local Similarity 31.6%; Pred. No. 2.6e-39;
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPNFIYTNNSLKDIINFYNNINDSKILSONKKNLVDTSYGNAEVSEGDVQLNP 62
Db 849 IPFOLSKYVDNQRLSTFTYEIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNDP 908
Qy 63 IFPPD---FKLGSSGDRGKVIYVTONENIVYNSMYESFSISFWIRINKVSNLP---GYT 116
Db 909 IDKQIQIOLFNESS-----KIEVLKNAIVYNSMYENFSTFWIRIPKYNFISLNNEYT 963
Qy 117 IDSVKNNSGWSIGIISNLFVFTLKQNESEQSINFSYDISNNAPGY-NKWFVTVTNMM 175
Db 964 IINCMMNSGKVSILNLTQDTQEIQRVWFYKYSQMINISDIYNRWIFVTITNR 1023
Qy 176 MGNKVIYNGKLIDTIKVKELTGINSKTTTFEINKIPDTGLTSDNINMWIRDFYIF 235
Db 1024 LNNKIYINGRLIDQKPIISNLGNHASNNIMFKDGCDDT-----HRYIWKYFNLF 1075
Qy 236 AKELDGKDINILFNSLOYTNVVDYNGNDLRYNKYMYNIDYLNRYMYANS----- 287
Db 1076 DKELNEKEIKDLYDNQNSGILKDFWGDYLOYDKPYMNLNLYDPNKYVDVNVGIRGYM 1135
Qy 288 ----RQIVFNTR-RNNDFNEGKIIIRKIRGTNTRVRGGDILYFDTINNKAYNLFM 342
Db 1136 LKGRPGSVMTNIVYNSLSYRGTRFIKKYASGNKNIVNNDRVINVVVNKEVRL-- 1193
Qy 343 KNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISG 402
Db 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 1246
Qy 403 ICSTGTFRFLGGDWYRHNLVPTVKOGNTASLLEST-----THWGFVPVSE 450
Db 1247 IGFIFGHQF-----NNIAKLVASNWNRYNQIERSSTRTLCGSWEFIPVDD 1289

RESULT 2
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919865
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
```

APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/480,604

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPND-01763

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 1296 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-915-136-28

Query Match 24.5%; Score 586.5; DB 4; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 2.6e-39;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

QY 3 IPFNFTNNSLLKDIINEFNNINSKILSLQNRKNTLVDSGYNAEVSSEGDVQLNP 62

DB 849 IPFQLSKYDQRLSTFTYEIKNIINTSIILNRYESNHLIDLSRYASKINIGSKVNFDP 908

QY 63 IPFED---FKLGSSGDRGKIVTQNEINIVNMYESFSFWIRNKWVSNLP---GYT 116

DB 909 IDKNOIQFNLESS-----KEVILKNAIVNSMYEFSFWIRIPKYNISLNEYT 963

QY 117 IIDSYKNSNGSIGIISNLFYFTLKQNSDSQSINFYDISNNAGY-NKWFVFTVNNM 175

DB 964 IINCENNSGKWSVLSNYGELIWTQTOEIKQVRVVKYSQMINISDYINRWIFVTINNR 1023

QY 176 MGNMKIYINGKLIDIKVKELTGINFSTITFEINKIPDAGLTISDSNINMWLRDPIYF 235

DB 1024 LNNSKIYINGRLIDQKPISNLGNHASNNIMFKLDGRDT-----HRYIWKYFNLF 1075

QY 236 AKELDKDIDILNFSLOQYTNVVDYKGNDRYNKYKYYMNIIDLYNRYMYANS----- 287  
DB 1076 DKELNEKEIKDLYDNQNSGILKDFWDYLDYKPYMLNLYDPNKYVDVNNVIGRGYMY 1135  
QY 288 ---RQIVFNTR-RNNDNFEGYKIIKIRIGTNTDTRVRGGDILYFDMTINNKAYNLFM 342  
DB 1136 LKPRGSGVMTNIVLNSSLYRGKFKIKKYASGNKNIVNRNDRVINNVYKKEVRL-- 1193  
QY 343 KNETMYADNHSTEDIYAIGLRBQTKDINDNIIQIOPMNTYYASQIFKSPNGENISG 402  
DB 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVYVVKSKNDQGITNKC-KMNLQDNNGND 1246  
QY 403 ICSIGTYRFRLLGGDWYRHNLYLPTVKQGNVYASLLEST-----THWGFVPVSE 450  
DB 1247 IGFIGFQF-----NNIAKLIVASNNWNRQIERSRRLGSGSWEFIPVDD 1289

#### RESULT 4

US-08-480-604A-23  
Sequence 23, Application US/08480604A  
Patent No. 5736139

#### GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32

#### CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104

#### COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30

#### CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,604A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/422,711

FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPND-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-480-604A-23

Query Match 23.3%; Score 557.5; DB 1; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

Qy 16 LKDIINEYFNINDSKILSLQNKNTLVTSGYNAEVSDEGVQLNPFPFDP---FKLGS 72  
Db 4 LLSTFEYIKNIINTSLNLRYESNHLIDLSRYASKINGSKVNFDPIDKQIQLFNLES 63  
Qy 73 SGEDRGKVIIVTQENIVYNSWYSPFSISFWIRKINKVSNLP---GYTIIDSVKYNSGWSI 129  
Db 64 S-----KIEVILKNAIVYNSWYENFSTFWIRPKYFNSISLNNEYTIINCNNNSGKWY 118  
Qy 130 GIISNFVFLTKQNEDESEQINSFYSDISNAPGY-NKWFVTVTNMGMNKIYINGKLI 188  
Db 119 SLNYGEIITWLTQTEIKQVYKYSQMINISDYINRWIFVITNNLNNSKIYINGRLI 178  
Qy 189 DTIKVKELTGINFSTKITFEINKIPDTGLTSDSDNNINWIRDFYIFAKELDGKINILF 248  
Db 179 DQKPISNLGNIHASNNIMFKLDCRDT-----HRYIWKYFNLFDKELNEKEIKDLY 230  
Qy 249 NSLQYTNVVDYNGNDLRYNKYEVYVNYDYLNRMYANS-----RQIVFNTFR- 295  
Db 231 DNQNSGILKDFWGDYLDQYDKPYTMLNLDPNKYDVNVNNGVYGLKPGSGVWTTNI 290  
Qy 296 RNNDFNEGKIIKIRKIRGNTNDRVRGGDILYFDMTINKKAYNLFMKNETWYADNHSTE 355  
Db 291 YLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRL-----ATNASQA 342  
Qy 356 DIVAIGLREQTKDINDNIIFQIOPMNTYVYASQIFKSNFNGENISGICSGTYFRFLGG 415  
Db 343 GVEKILSALEIPDVGNLSQVYVWKSNDQGITNKC-KMLQDNNGNDIGFIGHQF----- 397  
Qy 416 DWYRHNYLVPTVKQGNYSALLESTS-----THWGFVPVSE 450  
Db 398 -----NNIAKIVASNMWYNRQIERSRSLTGLCSWEFIPVDD 431

RESULT 5  
US-08-405-496A-23  
; Sequence 23, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; MEDIUM OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR-APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-405-496A-23

Query Match 23.3%; Score 557.5; DB 2; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;  
Qy 16 LKDIINEYFNINDSKILSLQNKNTLVTSGYNAEVSDEGVQLNPFPFDP---FKLGS 72  
Db 4 LLSTFEYIKNIINTSLNLRYESNHLIDLSRYASKINGSKVNFDPIDKQIQLFNLES 63  
Qy 73 SGEDRGKVIIVTQENIVYNSWYSPFSISFWIRKINKVSNLP---GYTIIDSVKYNSGWSI 129  
Db 64 S-----KIEVILKNAIVYNSWYENFSTFWIRPKYFNSISLNNEYTIINCNNNSGKWY 118  
Qy 130 GIISNFVFLTKQNEDESEQINSFYSDISNAPGY-NKWFVTVTNMGMNKIYINGKLI 188  
Db 119 SLNYGEIITWLTQTEIKQVYKYSQMINISDYINRWIFVITNNLNNSKIYINGRLI 178  
Qy 189 DTIKVKELTGINFSTKITFEINKIPDTGLTSDSDNNINWIRDFYIFAKELDGKINILF 248  
Db 179 DQKPISNLGNIHASNNIMFKLDCRDT-----HRYIWKYFNLFDKELNEKEIKDLY 230  
Qy 249 NSLQYTNVVDYNGNDLRYNKYEVYVNYDYLNRMYANS-----RQIVFNTFR- 295  
Db 231 DNQNSGILKDFWGDYLDQYDKPYTMLNLDPNKYDVNVNNGVYGLKPGSGVWTTNI 290  
Qy 296 RNNDFNEGKIIKIRKIRGNTNDRVRGGDILYFDMTINKKAYNLFMKNETWYADNHSTE 355  
Db 291 YLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRL-----ATNASQA 342  
Qy 356 DIVAIGLREQTKDINDNIIFQIOPMNTYVYASQIFKSNFNGENISGICSGTYFRFLGG 415  
Db 343 GVEKILSALEIPDVGNLSQVYVWKSNDQGITNKC-KMLQDNNGNDIGFIGHQF----- 397  
Qy 416 DWYRHNYLVPTVKQGNYSALLESTS-----THWGFVPVSE 450  
Db 398 -----NNIAKIVASNMWYNRQIERSRSLTGLCSWEFIPVDD 431

RESULT 6  
US-08-915-136-23  
; Sequence 23, Application US/08915136  
; Patent No. 6290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHIE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTIPOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MEDLEN & CARROLL, LLP  
;; STREET: 220 MONTGOMERY STREET, SUITE 2200  
;; CITY: SAN FRANCISCO  
;; STATE: CALIFORNIA  
;; COUNTRY: UNITED STATES OF AMERICA  
;; ZIP: 94104  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/915,136  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/480,604  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/405,496  
;; FILING DATE: 16-MAR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,154  
;; FILING DATE: 25-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/161,907  
;; FILING DATE: 02-DEC-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/985,321  
;; FILING DATE: 04-DEC-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/429,791  
;; FILING DATE: 31-OCT-1989  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: INGOLIA, DIANE E.  
;; REGISTRATION NUMBER: 40,027  
;; REFERENCE/DOCKET NUMBER: OPHD-01763  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 705-8410  
;; TELEFAX: (415) 397-8338  
;; INFORMATION FOR SEQ ID NO: 23:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 438 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-915-136-23

Query Match 23.3%; Score 557.5; DB 4; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;  
QY 16 LKDIINEYFNINDSKILSLQNKNTLVTSGYNAEVSSEGDVQLNPFFPD---FKLGS 72  
Db 4 LLSTFEYIKNIINTSLNRYESNHLDSRYASKINGSKVNTDPDKNQIQLFNLES 63  
QY 73 SGEDRGKVIYQENIVYNSMYESFISFIWINKWSNLP---GYTIIDSVKNNSGWSI 129  
Db 64 S-----KIEVLKNAIVYNSMYESFISFIWINKWSNLP---GYTIIDSVKNNSGWSI 118  
QY 130 GIISNLFVFTLKQNESEISNFSYDISNAPGY-NKFFFTVTNNMGNKVIYINGKLI 188  
Db 119 SLNVEIITWLOTQEQIKORVVKFSQMINISDYINRWTFVITNNRLNNSKIYINGRLI 178  
QY 189 DTIKVKELTGINSKTYTEINKIPDTGLITSDSNININWIRDFYIFAKELDGKDINILF 248  
Db 179 DQKPIISNLGINSNINIMPKLDCRDT-----HRYTIWKYFNLFKELNEKEIKOLY 230  
QY 249 NSLQYINVKYDNGNDLRNKEYEYMNIDYLNRYMYANS-----RQIVFNTR- 295  
Db 231 DNASGILKDFWGLDLYQIDYKPYIMLNLDYDPNRYVDNNVNGIRGMYLKGPRGVSNTINI 290

QY 296 RNNDFNEGYKIIIRKRGNTNDTRVRCGDILYFDMTNNKAYNLFMKNETMYADNHSTE 355  
Db 291 YLNSSLYRGTKFIIKKYASGNKDNIRNRNDRVYINNVYKNKEYRL-----ATNASQA 342  
QY 356 DIVAIGLREOTKDINDNIIFQIPMNNTYVYASQIFKSNFNGENISGICIGTYRFRLLGG 415  
Db 343 GYEKILSALEIFDVGNLQSVVYVYKSKNDQGITNKC-KNVLQDNNGNDIGFIFGHQF---- 397  
QY 416 DMRYRHNLYVTVKQGYNASLLESTST-----THWGFVFPVSE 450  
Db 398 -----NNIAKLVSANWYNRQIERSSRTILGCSWEFIPVDD 431  
RESULT 7  
US-08-480-604A-26  
; Sequence 26, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-480-604A-26

Query Match 23.3%; Score 557.5; DB 1; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEYNNINDSKILSONRKNLTVDTSYNAEVSSEGDVQLNPIFPD---FKLGS 72  
DB 28 LLSTFEYIKNIINTSIILNRYESNHLIDLRYASKINIGSKVNFDPDKNQIOIFNLES 87  
QY 73 SGEDRKVIVTQENIVNYSMEYSFISFWIRINKWVSNLP---GYTIIDSVKNNQWSI 129  
DB 88 S-----KIEVLKNAIVNYSMEYENSTFWIRIPKYNFNSISLNNEYTIINCMMENSGWKV 142  
QY 130 GIISNPLVTLKQNEDESEOSINFSYDINNAPGY-NKFFVTVTNMMGNKIINGKLI 188  
DB 143 SLNYGEIITLQDTQIKORVVKYSQMINISDYINRWIFVTIINRLNNSKIINGRLI 202  
QY 189 DTIKVRELGTINFSKTIITEINKIPDTGLITSDSNINKNWIRDFYIFAKELDGKDINILF 248  
DB 203 DQKPISNLGNIHASNIMEKLDGCRDT-----HRYIWIKYFNLFKELNEXEIKDLY 254  
QY 249 NSLOQYNNVYKDYNGNDLRYNKYKYYVNDILNRYMYANS-----RQIVFNTR- 295  
DB 255 DNQNSGILKDFWGDYLDYKPYMLNLYDPNKYVDVNVGIRGYMYLKGPRGSYMTINI 314  
QY 296 RNNDFNEGKIIKIRGNTNDTRVGGDILYFDMTINNKNAYNLFEMKNETMYADNHSTE 355  
DB 315 YLNSLYRGTKFIKKYASGKNDIVRNDRYINVVYKNEYRL-----ATNASQA 366  
QY 356 DIVAIGLREQTKDINDNIIFQIPMNTYYASQIFKSNFNGENISGCSIGTVRFRLG 415  
DB 367 GVEKILSALEIPDVGNLSQVVMYKSKNDQGITNKC-KMNLQDNNGNDIGFICFHQF---- 421  
QY 416 DWYRHNLYPTVKQGYASLLEST-----THWGFYVSE 450  
DB 422 -----NNIAKLVASNWNRYNQIERSSRTLCSSWEFIPVDD 455

## RESULT 8

US-08-405-496A-26  
Sequence 26, Application US/08405496A  
Patent No. 5919665  
GENERAL INFORMATION:  
APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-405-496A-26

Query Match 23.3%; Score 557.5; DB 2; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEYNNINDSKILSONRKNLTVDTSYNAEVSSEGDVQLNPIFPD---FKLGS 72  
DB 28 LLSTFEYIKNIINTSIILNRYESNHLIDLRYASKINIGSKVNFDPDKNQIOIFNLES 87  
QY 73 SGEDRKVIVTQENIVNYSMEYSFISFWIRINKWVSNLP---GYTIIDSVKNNQWSI 129  
DB 88 S-----KIEVLKNAIVNYSMEYENSTFWIRIPKYNFNSISLNNEYTIINCMMENSGWKV 142  
QY 130 GIISNPLVTLKQNEDESEOSINFSYDINNAPGY-NKFFVTVTNMMGNKIINGKLI 188  
DB 143 SLNYGEIITLQDTQIKORVVKYSQMINISDYINRWIFVTIINRLNNSKIINGRLI 202  
QY 189 DTIKVRELGTINFSKTIITEINKIPDTGLITSDSNINKNWIRDFYIFAKELDGKDINILF 248  
DB 203 DQKPISNLGNIHASNIMEKLDGCRDT-----HRYIWIKYFNLFKELNEXEIKDLY 254  
QY 249 NSLOQYNNVYKDYNGNDLRYNKYKYYVNDILNRYMYANS-----RQIVFNTR- 295  
DB 255 DNQNSGILKDFWGDYLDYKPYMLNLYDPNKYVDVNVGIRGYMYLKGPRGSYMTINI 314  
QY 296 RNNDFNEGKIIKIRGNTNDTRVGGDILYFDMTINNKNAYNLFEMKNETMYADNHSTE 355  
DB 315 YLNSLYRGTKFIKKYASGKNDIVRNDRYINVVYKNEYRL-----ATNASQA 366  
QY 356 DIVAIGLREQTKDINDNIIFQIPMNTYYASQIFKSNFNGENISGCSIGTVRFRLG 415  
DB 367 GVEKILSALEIPDVGNLSQVVMYKSKNDQGITNKC-KMNLQDNNGNDIGFICFHQF---- 421  
QY 416 DWYRHNLYPTVKQGYASLLEST-----THWGFYVSE 450  
DB 422 -----NNIAKLVASNWNRYNQIERSSRTLCSSWEFIPVDD 455

## RESULT 9

US-08-915-136-26  
Sequence 26, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA

;; COUNTRY: UNITED STATES OF AMERICA  
;; ZIP: 94104  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/915,136  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/480,604  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/405,496  
;; FILING DATE: 16-MAR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,154  
;; FILING DATE: 25-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/161,907  
;; FILING DATE: 02-DEC-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/985,321  
;; FILING DATE: 04-DEC-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/429,791  
;; FILING DATE: 31-OCT-1989  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: INGOLIA, DIANE E.  
;; REGISTRATION NUMBER: 40,027  
;; REFERENCE/DOCKET NUMBER: OPND-01763  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 705-8410  
;; TELEFAX: (415) 397-8338  
;; INFORMATION FOR SEQ ID NO: 26:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 462 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-915-136-26

Query Match 23.3%; Score 557.5; DB 4; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;  
QY 16 LKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSEGDVQINPFPPD---FKLGS 72  
DB 28 LLSTFETKNIINTSILNRYESNHLIDLSYASKINIGSKVNFDPIDKNOQLFNLES 87  
QY 73 SGDRGKVIVTQENIVNYSMTESFSFWIRINKVWSNLP---GYIIDSVKNNSGSI 129  
DB 88 S-----KIEVLKNAIVNYSMTESFSFWIRINKVWSNLP---GYIIDSVKNNSGSI 142  
QY 130 GIISNPLVFTLKQNEDEQSINFSYDINSNAPGY-NKWFVYVNNMGNKIYKGLI 188  
DB 143 SLNYGEIHTLODTEIKQVVKFSQMINSDYINRWIFVTIINRLNNSKIYINGLI 202  
QY 189 DTIKVKELTGINSFTIIFENKIPDTGLITSDSDINMWRIDYIFAKELDGKDINLF 248  
DB 203 DQKPIGNLNIHASNNMFKLDCRDT-----HRYIWKYFNLFKELNEKEIKDLY 254  
QY 249 NSLQYTNVVKDYGNDLRYNKEYYMVNIDYLNRYMYANS-----RQIVFVTR- 295  
DB 255 DNGSNGLKDWGLDQYDKRYFVLMNLDPNKYVVDVNNVIRGMYUKGPRGSMVTNI 314  
QY 296 RNNDFNEGYKIIIRKIRGTNDTRVGRGDILYFDMTINNKAYLMFMKNETMYADNHSTE 355  
DB 315 YLNSSLYRGFTFKIKKYGASGNDIVNRNDRVYINVVVKNEYRL-----ATNASQA 366  
QY 356 DIYAIGLRQTRDINDNIIFQIPMNNYNYASQIFKSNFNGENISGICSTGYRFLGG 415

DB 367 GVEKILSALEIPDVGNLSQVYVVMKSKNDQGITNKC-KMNLQDNNNGDIGFIFGHQF-----421  
QY 416 DWYRHNLYPTVKOGNTYASLES-----THWGFVPVSE 450  
DB 422 -----NNIAKLVASNWNRYNRQIERSRRTIGCSWEFIPYDD 455  
RESULT 10  
US-09-255-829-20  
;; Sequence 20, Application US/09255829  
;; Patent No. 6461617  
;; GENERAL INFORMATION:  
;; APPLICANT: Shone, Clifford Charles  
;; APPLICANT: Quinn, Conrad Padraig  
;; APPLICANT: Foster, Keith Alan  
;; TITLE OF INVENTION: Recombinant Toxin Fragments  
;; NUMBER OF SEQUENCES: 29  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.  
;; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
;; CITY: WASHINGTON  
;; STATE: DC  
;; COUNTRY: USA  
;; ZIP: 20005-3934  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/255,829  
;; FILING DATE: 23-FEB-1999  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB97/02273  
;; FILING DATE: 22-AUG-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/782,893  
;; FILING DATE: 27-DEC-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: ESMOND, ROBERT W.  
;; REGISTRATION NUMBER: 32,893  
;; REFERENCE/DOCKET NUMBER: 1581.0130002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-371-2600  
;; TELEFAX: 202-371-2540  
;; INFORMATION FOR SEQ ID NO: 20:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1169 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-255-829-20

Query Match 22.6%; Score 539.5; DB 4; Length 1169;  
Best Local Similarity 37.4%; Pred. No. 1.5e-35;  
Matches 130; Conservative 64; Mismatches 117; Indels 37; Gaps 12;  
QY 3 IFNFISFNNSSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSEGDVQINLP 62  
DB 836 MPFDUSTYNDTILFEMFNKNYSEILNLIINRYKNDNLIDLSYGAKVEYDGVDELN 895  
QY 63 IFPFDFKLGSSGEDRGKVIIVTQENIVNYSMTESFSFWIRINKV-----VSNL--PGYT 116  
DB 896 --KNQFKLTSSA--NSKIRVTQNQNIIFNSVFLDPSVFWIRIPKYNKDQNTIHNRYT 951  
QY 117 IDSVKNNSGWIGIISNPLVFTLKQNEDEQSINFSYDINSNAPGY-NKWFVYVNNM-175  
DB 952 IINCNNNSGWAISRGRIIWTLLIDINGKTSVFFENREDISEYINRFFVTIINN- 1010  
QY 176 MGNMKIYNGKLIDITKVKELTGINSFTIIFENKIPDTGLITSDSDINM-MWIRDFYI 234  
DB 1011 LNNAKIYINGKLESNTDKIDREVIANGELIFKLD-----GDIRTOFIWKKYFSI 1061





```

Db 574 NAYLANKWVFTTITNDRLSSANLYINGVLMSGAEITGLGAIRDNNITLKLDL----- 626
Qy 219 TSDSDNINMWIRDFYIFAKELDGKDINILFNLSQYTNVVKDYWGNDLRYNKEYYMYNI-- 276
Db 627 -CANNNGYVSIDKPRIFCKALNPKETKLYTSYLSITFLRDFWGNPLRYDTEYLLIPVAS 685
Qy 277 -----DYL---NRYWYANSRQIVENTRRNNDFNEGYKIIKKIRGNTN-DREV 321
Db 686 SSKDVQLKNITDYMYLTNAPSNGKLNYYRRLYN-----GLKFIKRYTPNNEIDSV 740
Qy 322 RGGDILYFDMTINNKA-----NLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 373
Db 741 KSGDFIKLYVSYNNNEHIVGPKDGNFNNLDRILRVGYNAPGIPLYKKWEAVK----- 794
Qy 374 IFQIQPMNTYYIASQIFKSNFNGENISGICSIGTYRFRLGSD-----WYRHNVL 423
Db 795 -----LRDLKTSYVQL--KLYDDKNAS-LGLVGTNGOIGNDPNDRDILIASNWY-FNEL 844
Qy 424 VPTVKQGNYSLESTSTHMGFVPVSE 450
Db 845 KDKI-----LGCDWYFVPTDE 860

```

Search completed: November 7, 2002, 14:47:30  
 Job time : 30 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:45:51 ; Search time 19 seconds  
(without alignments)  
341.491 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPFNIFSYNNLLKDI.....NYASLLESTHGWFPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586.5	24.5	1295	10 US-09-726-949A-1	Sequence 1, Appli
2	546	22.8	431	8 US-08-981-087A-1	Sequence 1, Appli
3	544.5	22.8	425	10 US-09-288-326-9	Sequence 9, Appli
4	250	10.5	144	8 US-08-981-087A-2	Sequence 2, Appli
5	239	10.0	144	8 US-08-981-087A-3	Sequence 3, Appli
6	145.5	6.1	1974	9 US-09-895-913A-12	Sequence 12, Appli
7	142.5	6.0	476	10 US-09-774-414-3	Sequence 3, Appli
8	133	5.6	932	10 US-09-815-242-5578	Sequence 5578, Ap
9	133	5.6	932	10 US-09-815-242-12438	Sequence 12438, A
10	129	5.4	841	10 US-09-815-242-5779	Sequence 5779, Ap
11	129	5.4	841	10 US-09-815-242-12751	Sequence 12751, A
12	128	5.4	1430	10 US-09-740-274-6	Sequence 6, Appli
13	121.5	5.1	990	12 US-10-047-678A-7	Sequence 7, Appli
14	121	5.1	483	8 US-08-834-666A-20	Sequence 20, Appli
15	118	4.9	993	10 US-09-815-242-5809	Sequence 5809, Ap
16	118	4.9	1002	10 US-09-815-242-12899	Sequence 12899, A
17	118	4.9	1002	10 US-09-815-242-13158	Sequence 13158, A
18	116.5	4.9	515	10 US-09-925-300-1282	Sequence 1282, Ap
19	116.5	4.9	996	10 US-09-815-242-5251	Sequence 5251, Ap

20	116.5	4.9	1009	10 US-09-815-242-12141	Sequence 12141, A
21	111.5	4.7	621	10 US-09-856-247A-2	Sequence 2, Appli
22	111.5	4.7	789	10 US-09-995-587A-1	Sequence 1, Appli
23	110.5	4.6	586	10 US-09-861-451A-32	Sequence 32, Appli
24	110	4.6	815	10 US-09-815-242-5106	Sequence 5106, Ap
25	110	4.6	824	10 US-09-866-582-34	Sequence 34, Appli
26	110	4.6	1116	10 US-09-790-318-2	Sequence 2, Appli
27	108	4.5	1437	10 US-09-801-368-354	Sequence 354, App
28	107	4.5	626	10 US-09-765-272-220	Sequence 220, App
29	107	4.5	869	10 US-09-815-242-10623	Sequence 10623, A
30	106.5	4.5	1781	9 US-09-995-749A-2	Sequence 2, Appli
31	106	4.4	691	8 US-08-834-666A-12	Sequence 12, Appli
32	106	4.4	691	8 US-08-834-666A-22	Sequence 22, Appli
33	104.5	4.4	555	8 US-08-808-031A-31	Sequence 31, Appli
34	104.5	4.4	586	8 US-08-808-031A-28	Sequence 28, Appli
35	103.5	4.3	584	9 US-09-995-749A-12	Sequence 12, Appli
36	103	4.3	1093	10 US-09-801-368-352	Sequence 352, App
37	102.5	4.3	1579	10 US-09-801-368-368	Sequence 368, App
38	101.5	4.2	476	10 US-09-733-524-15	Sequence 15, Appli
39	101.5	4.2	792	10 US-09-995-587A-11	Sequence 11, Appli
40	101	4.2	283	10 US-09-846-808-7	Sequence 7, Appli
41	101	4.2	283	12 US-10-059-964-26	Sequence 26, Appli
42	101	4.2	660	10 US-09-881-752A-22	Sequence 22, Appli
43	101	4.2	1143	10 US-09-924-154-14	Sequence 14, Appli
44	100.5	4.2	886	9 US-10-002-309B-2	Sequence 2, Appli
45	100.5	4.2	1031	10 US-09-815-242-10932	Sequence 10932, A

#### ALIGNMENTS

RESULT 1  
US-09-726-949A-1  
; Sequence 1, Application US/09726949A  
; Patent No. US20020137886A1  
; GENERAL INFORMATION:  
; APPLICANT: Allergan, Inc.  
; APPLICANT: Lin, Wei-Jen  
; APPLICANT: Aoki, Kei-Roger  
; APPLICANT: Steward, Lance E.  
; TITLE OF INVENTION: NEUROTOXINS WITH ENHANCED TARGET  
; TITLE OF INVENTION: SPECIFICITY  
; FILE REFERENCE: 36121-20002.00  
; CURRENT APPLICATION NUMBER: US/09/726,949A  
; CURRENT FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1295  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-09-726-949A-1

Query Match 24.5%; Score 586.5; DB 10; Length 1295;  
Best Local Similarity 31.6%; Pred. No. 4.3e-37;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
Qy 3 IPENFESYNNLLKDIINEYNNINDSKILQNRKNTIVDTSGYNAEYSEGDYQLNP 62  
Db 848 IPFQLSKVDNQRLLSTFTTEIKNIITSLNRYESNHLIDLSRYASKINIGSKYNEFP 907  
Qy 63 IPFPD---FKIGSSGDEGRKVIYQENINIVYNSMYEFSFWRINKWVSNLP---GYT 116  
Db 908 IDKNGIQLFNLESS-----KIEVLKNAIVNSMYENFSTSFWRIRPKYFNLSLNNEXT 962  
Qy 117 IIDSVKNSGSGSIGIISNFVITLQKNEDESDSEQSNFYSDISNNAPGY-NKWFVTVNNM 175  
Db 963 INCMENNSGKVSGLNTGEIITWLDQTEIKQRFVFKYSOMINISDIYNRWIVFTNNR 1022  
Qy 176 MGNMKIYNGLIIDTIKVKELTGINFSKITTFEINKIPDTGLITSDSDNNMIRDFYIF 235  
Db 1023 LNNKIYINGRLIQKPLISNLGNTHASNIMFKLDGCRDT-----HRYIWKYFNLF 1074





Db 228 DFWGDLQYDKPYMLNLPNKYVDVNNVIRGYMLKPGRGVMTNIIYLNSSLYRGT 287  
QY 306 KILIKIRGNTDTRVGGDILYEDMTINNKAYNLFKNETMYADNHSTEDIYAIGLREQ 365  
Db 288 KFIKKYASGNKNIVRNDRVINVVVRNKEIRL-----ATNASQAGVEKILSALE 339  
QY 366 TKDINDNIIQIOPMNNYYIASQIFKSNFNGENISGICISGTYFRGLGDMYRHNLYVP 425  
Db 340 IPDVGNLSQVVMKSKNDGIIKNC-KMNLQDNNGNDIGFIGHQF-----NNIAK 389  
QY 426 TVKQGNVASLLEST-----THWGFVPVSE 450  
Db 330 LVASNWNQIERSRRLGCSWEFIPVDD 418

RESULT 4  
US-08-981-087A-2  
; Sequence 2, Application US/08981087A  
; Patent No. US20020081304A1  
; GENERAL INFORMATION:  
; APPLICANT: Elmore, Michael J.  
; APPLICANT: Mauchline, Margaret L.  
; APPLICANT: Minton, Nigel P.  
; APPLICANT: Pasechnik, Vladimir A.  
; APPLICANT: Titball, Richard W.  
; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 No. US20020081304A1th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741

; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981,087A

; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01409

; FILING DATE: 12-JUN-1996  
; APPLICATION DATA:  
; APPLICATION NUMBER: GB 9511909.5

; FILING DATE: 12-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.

; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-688  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000

; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 144 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-981-087A-2

Query Match 10.5%; Score 250; DB 8; Length 144;  
Best Local Similarity 37.7%; Pred. No. 1.3e-12;  
Matches 55; Conservative 30; Mismatches 55; Indels 6; Gaps 4;

QY 9 SYTNLSLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEVSEGGVQLNPFPDF 68  
Db 1 SYTNKILILYFNKLYKKIKDNLDRYENKFDISGYGNSINISGDIYIYSTNRNOF 60

QY 69 KLGSSGDRGKVIQTQNTENIYVNSMYESFSISFIRINKWVS--NLPG-YTIDSVK-NN 124  
Db 61 GIYSS--KPEVNIAQNNDIYNGRYQNTSISFWRIPKYENKVNLNNEYTIIDCIRNN 118  
QY 125 SGWSIGLISFLVFLTKQNEDESEQSI 150  
Db 119 SGWKISLNYKIIWTILOTAGNNQKL 144

RESULT 5  
US-08-981-087A-3  
; Sequence 3, Application US/08981087A  
; Patent No. US20020081304A1  
; GENERAL INFORMATION:  
; APPLICANT: Elmore, Michael J.  
; APPLICANT: Mauchline, Margaret L.  
; APPLICANT: Minton, Nigel P.  
; APPLICANT: Pasechnik, Vladimir A.  
; APPLICANT: Titball, Richard W.  
; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 No. US20020081304A1th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741

; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981,087A

; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01409

; FILING DATE: 12-JUN-1996  
; APPLICATION DATA:  
; APPLICATION NUMBER: GB 9511909.5

; FILING DATE: 12-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.

; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-688  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000

; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 144 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-981-087A-3

Query Match 10.0%; Score 239; DB 8; Length 144;  
Best Local Similarity 37.0%; Pred. No. 9.1e-12;  
Matches 54; Conservative 25; Mismatches 57; Indels 10; Gaps 2;

QY 152 PSYDISNNAPGY-NKWFVTVNNMGMNKIYINGKLIDTIKVELTGINFSTIIFEIN 210  
Db 2 FNYTQMISIDSYINKWIFVTITNNRLGNSRIYINGNLDEKISINLGDHVSNNILFKIV 61

QY 211 KIPDTGLITSDSDNNMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRNKE 270  
Db 62 GCNDTRYV-----GIRFKVFOFTELKTELETLYSDPEPDSILKDFWGNLYNKR 112

QY 271 YMWYNDLYNRYMYANSRQIVFNTR 296

Db 113 YVLLNLRTRDKSITQNSFNINQQR 138

RESULT 6

US-09-895-913A-12

; Sequence 12, Application US/09895913A

; Patent No. US20020160456A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean Francois

; APPLICANT: Comen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in the

; TITLE OF INVENTION: Genome

; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 368

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12

; LENGTH: 1974

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-895-913A-12

Query Match 6.1%; Score 145.5; DB 9; Length 1974;

Best Local Similarity 21.2%; Pred. No. 0.003;

Matches 106; Conservative 61; Mismatches 143; Indels 191; Gaps 24;

QY 19 IINEYFNINDSKILSNKRNKTLVDTSGYNAEVEGDVQLNPFPDFKLGSSGEDRG 78

Db 60 IHSLFHTDADSKDQSVQRKQ-----FDFIPKTEVCVG 94

QY 79 KVIIVT--QNEIYVNSMYESESISFWIRINKWVSNLPGYTIIDSV-----KNSGWS 128

Db 95 VCYIAPYKNQDLIGSSAF-AWSLNF-----GATVVGTLILGSAQERANNNGS 141

QY 129 IGIISNFIIVF-----TLKQNEDEQSINFSYDISNNAFPGYKWFVTV 171

Db 142 IFWGNLILYLGHNATNIFLTNNVGNPNAGGGATINFNADETINADGLNTYTFQTV 201

QY 172 TNNMGNMKIY-----INGKLIDTK--VKELT--GINF-SKITPE-----INKI 212

Db 202 ALGLQTSASQHSWANFNLSKLSMEIKNSNFRDFTWGGFNFSGRITFTTFSGTWIN-- 259

QY 213 PDTGLITSDSNINWIRDFYIFAKELDGKDI-----NILFNSLQYT-----NVVKDY 260

Db 260 ---GATESGSSVVNNVANTDILFNSILGGIRYDLKANNLIENNSOMVIDSKNVNQSS 316

QY 261 WGNDRYKNEYVM-----NDYLN-----YMY 284

Db 317 LAGNVTFNNSRLSVKPNAAINIGDSOTQTALENASSLSFYNNSVANFPNGTAFNGVSYLN 376

QY 285 AN-SRQIVFTRNN-----DENEGKIIRKIRGTNTDTRVRGGDILY 328

Db 377 LNPNAQVSNQVNFNNVNYGYPFGKTFDFGNSARLI--NFKGTN-----F 424

QY 329 FDMTINNKAYL-----FMKNETMYADNHSTEDIYIAGLEQTKDINDNIIFQIQP 379

Db 425 NQATNLRAKNIHINFGQVSTFKQNSTMTNLAESSQAQFNALKVGEET---NFN----- 474

QY 380 MNNTYYASQIFKSNFNGENI 400

Db 475 LANSLL-----NFGNSV 488

RESULT 7

US-09-774-414-3

; Sequence 3, Application US/09774414

; Patent No. US20020102231A1

; GENERAL INFORMATION:

; APPLICANT: The Institute of Physical and Chemical Research

; TITLE OF INVENTION: Endonuclease

; FILE REFERENCE: PH-651

; CURRENT APPLICATION NUMBER: US/09/774,414

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 09/306,970

; PRIOR FILING DATE: 1999-05-07

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-774-414-3

Query Match 6.0%; Score 142.5; DB 10; Length 476;

Best Local Similarity 22.2%; Pred. No. 0.0089;

Matches 81; Conservative 58; Mismatches 105; Indels 121; Gaps 20;

QY 11 TNSLLKLD--IINEY-----FNNINDSKILSNKRNKTLVDTSGYNAEVS 53

Db 148 SNFYLMKYLINKYKYLDMLNMPNMYNMFNNYKGLNIK---TVLDLN--NNEFY 201

QY 54 E-----EGDVQLNPFPDFKLGSSGEDRGKVIYQENIYVNSMYESESISFWIRINK 107

Db 202 DYLSGLIEGDGYIGP-----GGTITNHANDVLTIFINKRNKNSILVEK 246

QY 108 WSNLPGYTIIDSVKNN---SGWSIGIISNEL---VFTLKQNE-DSEQSINFSYDISNN 159

Db 247 W-----MDILKQNPYVFNAFSINIKTLAKEIFTNINKLYSDYKIN--QINNH 294

QY 160 APGYNKWFFVTVN---NMGNMKIY-----INGKLI-----DTIKVKE----- 195

Db 295 IPYYN---YLKINKLPKIKMDIKNNYLAGTAADGSFLSSWYMPKDTLLFKNNRPSY 351

QY 196 -LTGINFSKITPEI-----NKIPDTGLITSDSNINWIRDFYIFAKE 238

Db 352 VISQVTERKELIYLOBSFDSLSNVKVKGNRKLDKFLTRTTDELMAKFIYFDKFLPL 411

QY 239 LOGKDINILNSQYTVNVKDYNGNDLRYKNEYMYNIDYLN-----YMYANSRQIV 291

Db 412 HDNQFNYI--KERENTFIKSYNWN---NRVGLVSEYINNINIKINDYDYNNKYINM 465

QY 292 ENTRR 296

Db 466 HNARK 470

RESULT 8

US-09-815-242-5578

; Sequence 5578, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/259,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5578  
; LENGTH: 932  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5578

Query Match 5.6%; Score 133; DB 10; Length 932;  
Best Local Similarity 17.6%; Pred. No. 0.011;  
Matches 86; Conservative 85; Mismatches 159; Indels 160; Gaps 19;

QY 6 NIFSYNNSLKDI-----INEFNNINDSKILSNKRLVDTSGYNA 50  
DB 250 NLYNAQGNIIAKGIYDSETSTTFTTYVDQYTNISGSFEQVAFAKRENATTDKTAIPM 309  
QY 51 EYSEEGD-VOLNPIPPDFKLGSSGDEGRKVIIVTONENIVNYSMEFSISFWIRNK-- 107  
DB 310 EYVLGNDKYSKNVIVDY-----GNKGQQLIS-STNYINN--EDLSRMTVYVYNQPK 358  
QY 108 -----WVSNLPGYTIIDSVKNNSGWSIGIISNPL-VFTLKONEDSEOSINFSYDISNN 159  
DB 359 KTYTKETFTVNLTKYKFNPDAKFKIYEVTNQNOFVDSFTPTSKLTDVTDKFKITYSND 418  
QY 160 APGVNKFVFTVNNMGNMKIYI-----NGKLIDTIKVK-----ELTG 198  
DB 419 ----NKTATVLLNGQSSDKQYIIQQVAPDNNSTNGKIDYILETQNGKSSWSNSYSN 474  
QY 199 INFSTIIFTEINKIPDTGLITSDS-----DNINMWIRDPYIFAKELDGKDN----- 245  
DB 475 VNGSSTANGDOKKYNLGDYVWEDTKDGKODANEKIGKGVYVILKDSNGKELDRTTDEN 534  
QY 246 -----ILFNSLQYTNVVKDYWGNDLRYNKE 270  
DB 535 GKYQFTGLNGTYSVEFTSLAGYPTTANAGTDDAIVSDGLTTGVKID--ADNMILDSG 592  
QY 271 YVMNIDYLNRYMYANSR-----QIVFNRNNN-----D 300  
DB 593 FYKTPKYSILGDYVWYDSNKGDKQDSTEKGKDKVKVTLNKEGEVIGTTKTDENGKYRFDN 652  
QY 301 FNEG-YKIIIRKIRGNT-----NDRVRGGDI-----LYFDMTINNKAYNLFMKNET 346  
DB 653 LDSGKYKVFIFEKPAGLTGTGTNTTEDDKADGGEVDVITDHDHDFLLDNGYY-----EET 708  
QY 347 MYADNHSTED 356  
DB 709 SDSDSDS 718

## RESULT 9

US-09-815-242-12438  
; Sequence 12438, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12438  
; LENGTH: 932  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12438

Query Match 5.6%; Score 133; DB 10; Length 932;  
Best Local Similarity 17.6%; Pred. No. 0.011;  
Matches 86; Conservative 85; Mismatches 159; Indels 160; Gaps 19;

QY 6 NIFSYNNSLKDI-----INEFNNINDSKILSNKRLVDTSGYNA 50  
DB 250 NLYNAQGNIIAKGIYDSETSTTFTTYVDQYTNISGSFEQVAFAKRENATTDKTAIPM 309  
QY 51 EYSEEGD-VOLNPIPPDFKLGSSGDEGRKVIIVTONENIVNYSMEFSISFWIRNK-- 107  
DB 310 EYVLGNDKYSKNVIVDY-----GNKGQQLIS-STNYINN--EDLSRMTVYVYNQPK 358  
QY 108 -----WVSNLPGYTIIDSVKNNSGWSIGIISNPL-VFTLKONEDSEOSINFSYDISNN 159  
DB 359 KTYTKETFTVNLTKYKFNPDAKFKIYEVTNQNOFVDSFTPTSKLTDVTDKFKITYSND 418  
QY 160 APGVNKFVFTVNNMGNMKIYI-----NGKLIDTIKVK-----ELTG 198  
DB 419 ----NKTATVLLNGQSSDKQYIIQQVAPDNNSTNGKIDYILETQNGKSSWSNSYSN 474  
QY 199 INFSTIIFTEINKIPDTGLITSDS-----DNINMWIRDPYIFAKELDGKDN----- 245  
DB 475 VNGSSTANGDOKKYNLGDYVWEDTKDGKODANEKIGKGVYVILKDSNGKELDRTTDEN 534  
QY 246 -----ILFNSLQYTNVVKDYWGNDLRYNKE 270  
DB 535 GKYQFTGLNGTYSVEFTSLAGYPTTANAGTDDAIVSDGLTTGVKID--ADNMILDSG 592  
QY 271 YVMNIDYLNRYMYANSR-----QIVFNRNNN-----D 300  
DB 593 FYKTPKYSILGDYVWYDSNKGDKQDSTEKGKDKVKVTLNKEGEVIGTTKTDENGKYRFDN 652  
QY 301 FNEG-YKIIIRKIRGNT-----NDRVRGGDI-----LYFDMTINNKAYNLFMKNET 346  
DB 653 LDSGKYKVFIFEKPAGLTGTGTNTTEDDKADGGEVDVITDHDHDFLLDNGYY-----EET 708  
QY 347 MYADNHSTED 356  
DB 709 SDSDSDS 718

## RESULT 10

US-09-815-242-5779  
; Sequence 5779, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5779  
LENGTH: 841  
TYPE: PRF  
ORGANISM: Staphylococcus aureus

US-09-815-242-5779

Query Match 5.4%; Score 129; DB 10; Length 841;  
Best Local Similarity 17.9%; Pred. No. 0.019;  
Matches 88; Conservative 81; Mismatches 160; Indels 162; Gaps 21;

QY 6 NIFSYTNNSLLKDI-----INEFNNINDSKILLSQNRKNTLVDTSGYNA 50  
DB 259 NLYNAQNIITAKGIYDSTNTTFTYTNVQDQTNVRSFEQVAFKRNATDTKAYKM 318  
QY 51 EVSEEGDVLNPIPFDFKLGSSGDRGKIVITQNNENIVNMYESFISFWIRINK--- 107  
DB 319 EVTLGNDYSEII-VDY-----GNKKAQPLIS-STNYINN---EDLSRMTAYVNPKN 368  
QY 108 -----WVSNLPGYTIIDSVKNNSGSI-----GLISNLFVTLKQNEDEQSNFSDIS 157  
DB 369 FYTKQTFVNTLTGYKFNPNKFNKFIYEVTDQNFVDSFPTDTSKLDKVDQ---FDVIYS 425  
QY 158 NNAPGYNKWFVTVTNMGMNKIYI-----NGKL---IDTIKVK-----EL 196  
DB 426 ND-----NKTATVLMKGTSSNKQYIIQQVAYPDNSSTDNGKIDYTLDTDKTKYSWSNSY 481  
QY 197 TGINFSTKITEINKIPDTGLTSDS-----DNINMWIRDFYIFAKELDGKDN----- 245  
DB 482 SNVNGSSANGDQKKYKNGLDYVWEDTNKQDQANEGKIGKVYVILKDSNGKELDRITTD 541  
QY 246 -----ILFNSLQYTNVVKDYWGNDRYN 268  
DB 542 ENGKYOFTGLSNGYSVEFSPAGYTTTANVGTDGDAVDSGLTTTCVIKD---ADNMILD 599  
QY 269 KEYVMNIDYLNRYMYANSR-----QIVFNTRRNN----- 299  
DB 600 SGFYKTPKYSGLDYVWYDSNKGQDQSTKGIKGVKVTQNEKGEVIGTTKTDENGKYRF 659  
QY 300 -DFNEG-VKILIKIRGNT-----NDTVRGSDILYFDMTINN---KAYNLFMKNE 345  
DB 660 DNDSDGKVKYFEKPAGLTQIVTNTWETDDKADGGEV---DVIITDHDFTLNGFEED 716  
QY 346 TMYADNHSSTED 356  
DB 717 TSDSDSDSDSD 727

RESULT 11  
US-09-815-242-12751  
Sequence 12751, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12751  
LENGTH: 841  
TYPE: PRF  
ORGANISM: Staphylococcus aureus

US-09-815-242-12751

Query Match 5.4%; Score 129; DB 10; Length 841;  
Best Local Similarity 17.9%; Pred. No. 0.019;  
Matches 88; Conservative 81; Mismatches 160; Indels 162; Gaps 21;

QY 6 NIFSYTNNSLLKDI-----INEFNNINDSKILLSQNRKNTLVDTSGYNA 50  
DB 259 NLYNAQNIITAKGIYDSTNTTFTYTNVQDQTNVRSFEQVAFKRNATDTKAYKM 318  
QY 51 EVSEEGDVLNPIPFDFKLGSSGDRGKIVITQNNENIVNMYESFISFWIRINK--- 107  
DB 319 EVTLGNDYSEII-VDY-----GNKKAQPLIS-STNYINN---EDLSRMTAYVNPKN 368  
QY 108 -----WVSNLPGYTIIDSVKNNSGSI-----GLISNLFVTLKQNEDEQSNFSDIS 157  
DB 369 FYTKQTFVNTLTGYKFNPNKFNKFIYEVTDQNFVDSFPTDTSKLDKVDQ---FDVIYS 425  
QY 158 NNAPGYNKWFVTVTNMGMNKIYI-----NGKL---IDTIKVK-----EL 196  
DB 426 ND-----NKTATVLMKGTSSNKQYIIQQVAYPDNSSTDNGKIDYTLDTDKTKYSWSNSY 481  
QY 197 TGINFSTKITEINKIPDTGLTSDS-----DNINMWIRDFYIFAKELDGKDN----- 245  
DB 482 SNVNGSSANGDQKKYKNGLDYVWEDTNKQDQANEGKIGKVYVILKDSNGKELDRITTD 541  
QY 246 -----ILFNSLQYTNVVKDYWGNDRYN 268  
DB 542 ENGKYOFTGLSNGYSVEFSPAGYTTTANVGTDGDAVDSGLTTTCVIKD---ADNMILD 599  
QY 269 KEYVMNIDYLNRYMYANSR-----QIVFNTRRNN----- 299  
DB 600 SGFYKTPKYSGLDYVWYDSNKGQDQSTKGIKGVKVTQNEKGEVIGTTKTDENGKYRF 659

Qy 300 -DFNEG-YKIIKIRNT-----NDRVRGGDILYFDMTINN-----KAYNLFMKE 345  
Db 660 DNLDSGKYKFIKXEPAGLTQTNTTDDKADGGV--DVTITDHDFTLDDNGYFEE 716  
Qy 346 TMVADNHSTED 356  
Db 717 TSDSDSDSDSD 727

RESULT 12  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 5.4%; Score 128; DB 10; Length 1430;  
Best Local Similarity 20.6%; Pred. No. 0.044;  
Matches 117; Conservative 73; Mismatches 188; Indels 190; Gaps 30;

Qy 10 YTNNSLL-KDINEYFNND--SKILSLQNRKNT---LVDFSGYNA-----EYSEEG 56  
Db 850 YESSALDSQLYEGSFNFQDFYTKDSYTNKKIAQNVQLFKSGVTSFEMAPQYVSSD 909  
Qy 57 DVQNLPTFPDFKLGSGGDRGKVVIVTONENIVNMSYESFS-----ISFWIR 104  
Db 910 GSFLDSI-----IQNGYAFEDRYDLAMSKNNK--YGSQODMINAVKALHKSQIQVIADW-- 961  
Qy 105 INKAVSNLPGYTIIDSVK--NNSGWSIGIISNLFVTLTKQEDSE--QSINFSDISNAPG 162  
Db 962 VPDQIYNLPGEVTVATRVNDYG-----EYKDSIEINTLYAANKSNGKD 1007  
Qy 163 YNKFFVTVTNMNG-----NMKIYNGKID-----TIKVELTGINF----- 201  
Db 1008 YQAKYGGAFLSLAAPYSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILRGVGYVL 1067  
Qy 202 ---SKTITFEI-----NKPIDTGLI-----TSDSDNINMWIRD---F 232  
Db 1068 KDNASDKYFELKGNQTVLPQMTNKEASTGFVNDGNGMFTYSGYQAKNSFVQDAGKMW 1127  
Qy 233 YIF-----AKELDGKJNILFNSLOY-----TNVVKDYWGN-DLRYNKEYMYN 275  
Db 1128 YFDFNNGHMVYGLQALNGEVQYFLSNGVQVRESFLENADGSKNYFGLGNRYSNGYVSD 1187  
Qy 276 IDYLNRYMYANSRQIV-FNTRRNNDP--NEGKIIKIRGNTNDRVRGGDILYFDMT 332  
Db 1188 NDSKRWYFDAGVMAVGLKTINGNTQYFDQDQYGVKAWITGSDGKKR-----YFDDG 1240

Qy 333 INKAYNLFM--KNETMYADNHSTEDIYAIGLRE-----QTKDINDN--- 372  
Db 1241 SGMAVNRFAANDKNGDWYLN--SDGIALVGVQTINGKTYTQGDKQIKGRIITDNKGL 1298  
Qy 373 -----IIFQIOPMNTYVYASQ-----IFKSN-----F 395  
Db 1299 KYFLANGSELARNIFATDSQNNWYFYGSDGVAVTGSQTAGKKLYFASDGKQVKGSEFTY 1358  
Qy 396 NGE-----NISGICSGTYFRRLGGDWY 418  
Db 1359 NGKVHYVHADSGELQVNRFEADKDGWY 1386

RESULT 13  
US-10-047-676A-7  
; Sequence 7, Application US/10047676A  
; Patent No. US20020123105A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Fengxia  
; APPLICANT: Caulfield, Page W.  
; APPLICANT: Chen, Ping W.  
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS  
; FILE REFERENCE: UAB-17403/22  
; CURRENT APPLICATION NUMBER: US/10/047,676A  
; CURRENT FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 09/627,376  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 990  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-047-676A-7

Query Match 5.1%; Score 121.5; DB 12; Length 990;  
Best Local Similarity 21.9%; Pred. No. 0.087;  
Matches 107; Conservative 86; Mismatches 170; Indels 125; Gaps 29;

Qy 12 NNSLL-----KDINEYFN-NIND--SKILSLQNRKNTLVDTSGY-----NAEVS- 53  
Db 254 NNSLELLEKLEIKAMITQDYTRNIGSGNSILALENKSQIVKANAYLRVLDYDHAELKL 313  
Qy 54 ---BEGDVQ--LNPFFPDFKLGSGGDRGKVVIVTONENIVNMSYESFSISFWIRNKW 108  
Db 314 AQHTKSLQNLILKVLSSF-----SSAVNSQKEIKNYHEKFIARYGYEQ-L-VPLQLLLNS- 366  
Qy 109 VSNL---PGYTIID-SYKNNSGWSIGIISNLFVTLTKQ-----NEDGEQSTNFSY 154  
Db 367 TSGLGFPKGSQTEVSKQNNEDSKNOKIIEFLQKFEKALRDGKEIILSDDDLKDLNFDI 426  
Qy 155 DISNAPGYNKKFFVTV-----TNMMGNMKIYINGKLIDTI-----VKVELTGIN 200  
Db 427 EQQISGELYCYFNFKSKKLEVSISLGVSQMLGNTFGRFHSHKLPNTIVTKNVNKTKEIFTEA 486  
Qy 201 FSKTITEINKIPDTG-----LITSDDNINMWIRDFYIFAKELDGKDINILFNSLOYTN 255  
Db 487 IPNTIITQLNEVPVFGGGMIMLSLKSQLELRN-YTTKKEMSINDIYVTRATSEEL-- 543  
Qy 256 VVKDYWGNDRYRNEY-----YMNVDILNRYMYAN-SRQIVFENTERNNNDFN----- 302  
Db 544 ---YF-----YSKKYEKRVIFVMN-----NMENYINGSKLLRELLEVSNSDFQNTIPTL 590  
Qy 303 ---BQYK-----IIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMK--NETMY 348  
Db 591 GSLDSYNHVPATYIKDIIIIKPEWNTKSEAKTLDLKNLWLTNNVPPFRMKYIDQIIY 650  
Qy 349 ADNHSTEDIYAI--GLRE-----QTKDI-----NDNIFIQ-IOPMNT-----YYASQ 389  
Db 651 LDLSTIDLMLFQSIKKHSFIQLLDVHSVCTINDTEILELVVPEFTRSDVNAHQIYHAQN 710  
Qy 390 IFKSNFNG 397



```

QY 171 VTNN--MMGNKTYING--KLIDTIVKVELTGINSKTIITFEINKIPDTGLITSDDNIN 226
Db 191 VTNNITFSNLAEKINGEPEINDITKVYVIHNLVSKELIYSTIRPP-----LSYSDNIN 245
QY 227 MWIRDFYI---FAKELDGKDINILFNSLOYTNVVDYWGNDLRN-----KEY 271
Db 246 YILNKLSLHNDDEFVKTI--REIQKLIILAVEKTEI--GFEELYKDIIHHMKALFKCKNY 300
QY 272 YMVN--IDYLNRYMY-----ANSROIVENTRRNNNDENE-----GYKIIIK 310
Db 301 LQIDTKIDMINNYLHQDIATNISEAYALLLSRNNIGFTDLKVLHNRFEIKYGEOLV- 359
QY 311 RIRGTNDTRVGGDILYFDMTINNKAYNLFMKNETMYADNHESTEDIYAIGLREOTKDIN 370
Db 360 NIKDLISDITGFGTTIFQEBETDCNNI--VMLKQKFLHALRNNDEIV--INEKDVESLIN 415
QY 371 DNIIFQIQPMNNTYYIASQIFKSNFNGENISGICSIGTYFRGLGGDWYRHNVLVPTV--- 427
Db 416 DNEINHYPMSADVYAEVLGRFYNOYNELIVPLTASFNAGATGFRFHLIDPTELA 475
QY 428 ---KQGN 432
Db 476 KLEHEKGY 484

```

Search completed: November 7, 2002, 14:53:54  
Job time : 24 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:43:16 ; Search time 286 Seconds  
(without alignments)  
1014,440 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTPFNFTSNNLLKDI.....NYASLLESTTHWGFVPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

```
1: /cgn2_6/ptodata/1/paa/pctus_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2391	100.0	450	20	US-09-611-419A-10
2	2391	100.0	450	23	US-09-910-186A-10
3	2386	99.8	1291	1	PCT-US97-15394-60
4	2386	99.8	1291	11	US-08-704-159-60
5	2386	99.8	1291	13	US-08-954-302-1
6	2386	99.8	1291	21	US-09-791-537-13524

7	2386	99.8	1291	21	US-09-791-537-95055
8	2386	99.8	1291	26	US-10-205-516-6
9	2386	99.8	1291	26	US-10-271-012-60
10	2386	99.8	1301	26	US-10-205-516-20
11	2317	96.9	462	1	PCT-US97-15394-62
12	2317	96.9	462	11	US-08-704-159-62
13	2317	96.9	462	26	US-10-271-012-62
14	1824	76.3	1285	21	US-09-791-537-59796
15	1823	76.2	1285	21	US-09-791-537-18595
16	882.5	36.9	1280	21	US-09-791-537-75966
17	882.5	36.9	1280	21	US-09-791-537-112792
18	867.5	36.3	451	20	US-09-611-419A-12
19	867.5	36.3	1275	21	US-09-791-537-33117
20	867.5	36.3	1276	1	PCT-US97-15394-66
21	867.5	36.3	1276	11	US-08-704-159-66
22	867.5	36.3	1276	19	US-09-547-188-8
23	867.5	36.3	1276	21	US-09-791-537-13527
24	867.5	36.3	1276	21	US-09-791-537-92962
25	867.5	36.3	1276	26	US-10-205-516-8
26	867.5	36.3	1276	26	US-10-271-012-66
27	867.5	36.3	1286	26	US-10-205-516-22
28	864.5	36.2	451	23	US-09-910-186A-12
29	801.5	33.5	451	1	PCT-US97-15394-68
30	801.5	33.5	451	11	US-08-704-159-68
31	801.5	33.5	451	26	US-10-271-012-68
32	604.5	25.3	1297	1	PCT-US97-15394-77
33	604.5	25.3	1297	11	US-08-704-159-77
34	604.5	25.3	1297	19	US-09-547-188-14
35	604.5	25.3	1297	21	US-09-791-537-14255
36	604.5	25.3	1297	21	US-09-791-537-132350
37	604.5	25.3	1297	26	US-10-205-516-14
38	604.5	25.3	1297	26	US-10-271-012-77
39	604.5	25.3	1307	26	US-10-205-516-28
40	602.5	25.2	1252	26	US-10-205-516-10
41	602.5	25.2	1262	26	US-10-205-516-24
42	595.5	24.9	1250	1	PCT-US97-15394-50
43	595.5	24.9	1250	11	US-08-704-159-50
44	595.5	24.9	1250	26	US-10-271-012-50
45	594.5	24.9	449	23	US-09-910-186A-14

#### ALIGNMENTS

RESULT 1  
US-09-611-419A-10  
; Sequence 10, Application US/09611419A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Leonard A.  
; APPLICANT: Byrne, Michael P.  
; APPLICANT: Middlebrook, John L.  
; APPLICANT: Lapenotiere, Hugh  
; APPLICANT: Clayton, Michael A.  
; APPLICANT: Brown, Douglas R.  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; FILE REFERENCE: A33626 067252.0105  
; CURRENT APPLICATION NUMBER: US/09/611,419A  
; PRIORITY FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: PCT/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/146,192  
; PRIOR FILING DATE: 1999-07-29



```
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Encoded polypeptide of a synthetic construct based
; on BoNTC HC
US-09-611-419A-10
```

```
Query Match 100.0%; Score 2391; DB 20; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.8e-212;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTIPNFSTNNSLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
Db 1 MTIPNFSTNNSLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
Qy 61 NPFPDFKLGSGEDRGKIVYTONENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
Db 61 NPFPDFKLGSGEDRGKIVYTONENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
Qy 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDSSNAPGYNKWFVYVTNNMGNMK 180
Db 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDSSNAPGYNKWFVYVTNNMGNMK 180
Qy 181 IYINGKLIDITKVELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 240
Db 181 IYINGKLIDITKVELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 240
Qy 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYMVNIDYLNRYMYANSRQIVENTRRNND 300
Db 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYMVNIDYLNRYMYANSRQIVENTRRNND 300
Qy 301 FNEGKIIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNSTDIEDIYAI 360
Db 301 FNEGKIIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNSTDIEDIYAI 360
Qy 361 GLREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGSDWYRH 420
Db 361 GLREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGSDWYRH 420
Qy 421 NYLPTVKQGNYSALLESTSTHMGFVPVSE 450
Db 421 NYLPTVKQGNYSALLESTSTHMGFVPVSE 450
```

```
RESULT 2
US-09-910-186A-10
; Sequence 10, Application US/09910186A
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Material Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; FILE REFERENCE: A33626-A 067252.0107
; CURRENT APPLICATION NUMBER: US/09/910,186A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
```

```
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; OTHER INFORMATION: sequence
US-09-910-186A-10

Query Match 100.0%; Score 2391; DB 23; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.8e-212;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTIPNFSTNNSLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
Db 1 MTIPNFSTNNSLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
Qy 61 NPFPDFKLGSGEDRGKIVYTONENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
Db 61 NPFPDFKLGSGEDRGKIVYTONENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
Qy 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDSSNAPGYNKWFVYVTNNMGNMK 180
Db 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDSSNAPGYNKWFVYVTNNMGNMK 180
Qy 181 IYINGKLIDITKVELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 240
Db 181 IYINGKLIDITKVELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 240
Qy 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYMVNIDYLNRYMYANSRQIVENTRRNND 300
Db 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYMVNIDYLNRYMYANSRQIVENTRRNND 300
Qy 301 FNEGKIIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNSTDIEDIYAI 360
Db 301 FNEGKIIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNSTDIEDIYAI 360
Qy 361 GLREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGSDWYRH 420
Db 361 GLREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGSDWYRH 420
Qy 421 NYLPTVKQGNYSALLESTSTHMGFVPVSE 450
Db 421 NYLPTVKQGNYSALLESTSTHMGFVPVSE 450
```

```
RESULT 3
PCT-US97-15394-60
; Sequence 60, Application PC/TUS9715394
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: PCT/US97/15394  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US97-15394-60

Query Match 99.8%; Score 2386; DB 1; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFNIFSTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEYSEGGVQLN 61  
DB 843 TIFNIFSTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEYSEGGVQLN 902  
QY 62 PIFPDKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRINKVWSNLPGYTIIDSV 121  
DB 903 PIFPDKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRINKVWSNLPGYTIIDSV 962  
QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 1022  
QY 182 YINGKLIDITIKVKELTGINFSTIIFKIPDGLITSDSDNINMWIRDFYIFAKELDG 241  
DB 1023 YINGKLIDITIKVKELTGINFSTIIFKIPDGLITSDSDNINMWIRDFYIFAKELDG 1082  
QY 242 KDINILFNSLQTVNVKDYWGNDLRNKEYYVWIDYLNRYMYANSQIVFTRNNND 301  
DB 1083 KDINILFNSLQTVNVKDYWGNDLRNKEYYVWIDYLNRYMYANSQIVFTRNNND 1142  
QY 302 NEGYKIIIRKRGNTNDRVRGGDILYFDMTINKAYNLFKMETMADNHSTEDIYAIG 361  
DB 1143 NEGYKIIIRKRGNTNDRVRGGDILYFDMTINKAYNLFKMETMADNHSTEDIYAIG 1202  
QY 362 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 421  
DB 1203 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 1262

RESULT 4  
US-08-704-159-60  
Sequence 60, Application US/08704159  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine for Clostridium  
TITLE OF INVENTION: Botulinum Neurotoxin  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,159  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-704-159-60

Query Match 99.8%; Score 2386; DB 1; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFNIFSTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEYSEGGVQLN 61  
DB 843 TIFNIFSTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEYSEGGVQLN 902  
QY 62 PIFPDKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRINKVWSNLPGYTIIDSV 121  
DB 903 PIFPDKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRINKVWSNLPGYTIIDSV 962  
QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 1022  
QY 182 YINGKLIDITIKVKELTGINFSTIIFKIPDGLITSDSDNINMWIRDFYIFAKELDG 241  
DB 1023 YINGKLIDITIKVKELTGINFSTIIFKIPDGLITSDSDNINMWIRDFYIFAKELDG 1082  
QY 242 KDINILFNSLQTVNVKDYWGNDLRNKEYYVWIDYLNRYMYANSQIVFTRNNND 301  
DB 1083 KDINILFNSLQTVNVKDYWGNDLRNKEYYVWIDYLNRYMYANSQIVFTRNNND 1142  
QY 302 NEGYKIIIRKRGNTNDRVRGGDILYFDMTINKAYNLFKMETMADNHSTEDIYAIG 361  
DB 1143 NEGYKIIIRKRGNTNDRVRGGDILYFDMTINKAYNLFKMETMADNHSTEDIYAIG 1202  
QY 362 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 421  
DB 1203 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 1262  
QY 422 YLVPVTKQGNYSALLESSTHWGFVPVSE 450  
DB 1263 YLVPVTKQGNYSALLESSTHWGFVPVSE 1291

RESULT 5  
US-08-954-302-1  
Sequence 1, Application US/08954302  
GENERAL INFORMATION:  
APPLICANT: Lance Simpson, Nikita Kiyatkin,  
APPLICANT: Andrew Maksymowich  
TITLE OF INVENTION: Compositions and Methods for Systemic  
TITLE OF INVENTION: Delivery of Oral Vaccines and Therapeutic Agents  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jane Massey Licata, Esq.  
STREET: 66 E. Main Street  
CITY: Marlton  
STATE: NJ

COUNTRY: USA  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/954,302  
FILING DATE: herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane Massey Licata  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: JEFF-0164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 779-2400  
TELEFAX: (609) 810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
US-08-954-302-1

Query Match 99.8%; Score 2386; DB 13; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIPFNISYTNNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 61  
DB 843 TIPFNISYTNNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 902

QY 62 PIPFPDKLGSSGEGDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
DB 903 PIPFPDKLGSSGEGDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 1022

QY 182 YINGKLIDTIKVKELTGINFSKITTEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 241  
DB 1023 YINGKLIDTIKVKELTGINFSKITTEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 1082

QY 242 KDNILFNSLQYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 301  
DB 1083 KDNILFNSLQYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 1142

QY 302 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361  
DB 1143 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202

QY 362 LREOTKDINDIIFQIPMNNTYVYASQIPKSNFNGENISGICSGTYFRGLGGDWYRHN 421  
DB 1203 LREOTKDINDIIFQIPMNNTYVYASQIPKSNFNGENISGICSGTYFRGLGGDWYRHN 1262

QY 422 YLVPVTKQGNYSALLESTSTHWGFVPVSE 450  
DB 1263 YLVPVTKQGNYSALLESTSTHWGFVPVSE 1291

RESULT 6  
US-09-791-537-13524  
; Sequence 13524, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB

; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13524  
; LENGTH: 1291  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-09-791-537-13524

Query Match 99.8%; Score 2386; DB 21; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIPFNISYTNNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 61  
DB 843 TIPFNISYTNNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 902

QY 62 PIPFPDKLGSSGEGDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
DB 903 PIPFPDKLGSSGEGDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 1022

QY 182 YINGKLIDTIKVKELTGINFSKITTEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 241  
DB 1023 YINGKLIDTIKVKELTGINFSKITTEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 1082

QY 242 KDNILFNSLQYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 301  
DB 1083 KDNILFNSLQYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 1142

QY 302 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361  
DB 1143 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202

QY 362 LREOTKDINDIIFQIPMNNTYVYASQIPKSNFNGENISGICSGTYFRGLGGDWYRHN 421  
DB 1203 LREOTKDINDIIFQIPMNNTYVYASQIPKSNFNGENISGICSGTYFRGLGGDWYRHN 1262

QY 422 YLVPVTKQGNYSALLESTSTHWGFVPVSE 450  
DB 1263 YLVPVTKQGNYSALLESTSTHWGFVPVSE 1291

RESULT 7  
US-09-791-537-95055  
; Sequence 95055, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 95055  
; LENGTH: 1291  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum phage  
US-09-791-537-95055

Query Match 99.8%; Score 2386; DB 21; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 TTPFNIFSYTNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 61
|
|
|
Db 843 TTPFNIFSYTNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 902
|
|
|
QY 62 PTFPFDFKLGSGGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 121
|
|
|
Db 903 PTFPFDFKLGSGGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 962
|
|
|
QY 122 KNSGWSIGIISNFIYFTLKQNEDEQSIYSDISNAPGNKWFVTVTNMMGNMKI 181
|
|
|
Db 963 KNSGWSIGIISNFIYFTLKQNEDEQSIYSDISNAPGNKWFVTVTNMMGNMKI 1022
|
|
|
QY 182 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241
|
|
|
Db 1023 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
|
|
|
QY 242 KDINILFNSLQTVNVKDYWGNDLRYNKKEYTVMWIDYLNRYMYANSQIIVENTRRNNDF 301
|
|
|
Db 1083 KDINILFNSLQTVNVKDYWGNDLRYNKKEYTVMWIDYLNRYMYANSQIIVENTRRNNDF 1142
|
|
|
QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLMKNETYADNHSTEDIYAIG 361
|
|
|
Db 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLMKNETYADNHSTEDIYAIG 1202
|
|
|
QY 362 LREQTKDINDNTIIFOQPMNNTIYVYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 421
|
|
|
Db 1203 LREQTKDINDNTIIFOQPMNNTIYVYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 1262
|
|
|
QY 422 YLVPTVKQGNYSALLESTSTHGWFPVSE 450
|
|
|
Db 1263 YLVPTVKQGNYSALLESTSTHGWFPVSE 1291
|
|
|
RESULT 8
US-10-205-516-6
; Sequence 6, Application US/10205516
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; FILE REFERENCE: J2bt1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 6
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-6

Query Match 99.8%; Score 2386; DB 26; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.9e-211;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTPFNIFSYTNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 61
|
|
|
Db 843 TTPFNIFSYTNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 902
|
|
|
QY 62 PTFPFDFKLGSGGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 121
|
|
|
Db 903 PTFPFDFKLGSGGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 962
|
|
|
QY 122 KNSGWSIGIISNFIYFTLKQNEDEQSIYSDISNAPGNKWFVTVTNMMGNMKI 181
|
|
|
Db 963 KNSGWSIGIISNFIYFTLKQNEDEQSIYSDISNAPGNKWFVTVTNMMGNMKI 1022
|
|
|
QY 182 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241
|
|
|
Db 1023 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
|
|
|
QY 242 KDINILFNSLQTVNVKDYWGNDLRYNKKEYTVMWIDYLNRYMYANSQIIVENTRRNNDF 301
|
|
|

```

```

Db 1083 KDINILFNSLQTVNVKDYWGNDLRYNKKEYTVMWIDYLNRYMYANSQIIVENTRRNNDF 1142
|
|
|
QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLMKNETYADNHSTEDIYAIG 361
|
|
|
Db 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLMKNETYADNHSTEDIYAIG 1202
|
|
|
QY 362 LREQTKDINDNTIIFOQPMNNTIYVYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 421
|
|
|
Db 1203 LREQTKDINDNTIIFOQPMNNTIYVYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 1262
|
|
|
QY 422 YLVPTVKQGNYSALLESTSTHGWFPVSE 450
|
|
|
Db 1263 YLVPTVKQGNYSALLESTSTHGWFPVSE 1291
|
|
|
RESULT 9
US-10-271-012-60
; Sequence 60, Application US/10271012
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-271-012-60

Query Match 99.8%; Score 2386; DB 26; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.9e-211;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTPFNIFSYTNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 61
|
|
|
Db 843 TTPFNIFSYTNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 902
|
|
|
QY 62 PTFPFDFKLGSGGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 121
|
|
|
Db 903 PTFPFDFKLGSGGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 962
|
|
|
QY 122 KNSGWSIGIISNFIYFTLKQNEDEQSIYSDISNAPGNKWFVTVTNMMGNMKI 181
|
|
|

```

Db 963 KNSGWSGIISNLFVTLKQNEDESEISINFSYDISNAPGYNKFFVTVTNNMGMNMI 1022  
Qy 182 YNGKLIDTIKVKELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 241  
Db 1023 YNGKLIDTIKVKELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 1082  
Qy 242 KDINILFNSLOYTNVVDYWGNDLRYNKEYYVMVNDYLNRYMYANSRQIVFNTRRNNDF 301  
Db 1083 KDINILFNSLOYTNVVDYWGNDLRYNKEYYVMVNDYLNRYMYANSRQIVFNTRRNNDF 1142  
Qy 302 NEGKIIIRKIRGNNTDTRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 361  
Db 1143 NEGKIIIRKIRGNNTDTRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 1202  
Qy 362 LREQTKDINDNIIFQIOPMNNYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHN 421  
Db 1203 LREQTKDINDNIIFQIOPMNNYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHN 1262  
Qy 422 YLVPVKOGNYASLLESTHGWGFPVSE 450  
Db 1263 YLVPVKOGNYASLLESTHGWGFPVSE 1291  
RESULT 10  
US-10-205-516-20  
; Sequence 20, Application US/10205516  
; GENERAL INFORMATION:  
; APPLICANT: Zhong, Jun  
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
; FILE REFERENCE: Jb7bx1  
; CURRENT APPLICATION NUMBER: US/10/205,516  
; CURRENT FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1301  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-205-516-20  
Query Match 99.8%; Score 2386; DB 26; Length 1301;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 TIFPFIYSYTNNSLLKDIIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61  
Db 847 TIFPFIYSYTNNSLLKDIIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 906  
Qy 62 PIFPFDKLGSGGDRGKVIYQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
Db 907 PIFPFDKLGSGGDRGKVIYQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 966  
Qy 122 KNSGWSGIISNLFVTLKQNEDESEISINFSYDISNAPGYNKFFVTVTNNMGMNMI 181  
Db 967 KNSGWSGIISNLFVTLKQNEDESEISINFSYDISNAPGYNKFFVTVTNNMGMNMI 1026  
Qy 182 YNGKLIDTIKVKELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 241  
Db 1027 YNGKLIDTIKVKELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 1086  
Qy 242 KDINILFNSLOYTNVVDYWGNDLRYNKEYYVMVNDYLNRYMYANSRQIVFNTRRNNDF 301  
Db 1087 KDINILFNSLOYTNVVDYWGNDLRYNKEYYVMVNDYLNRYMYANSRQIVFNTRRNNDF 1146  
Qy 302 NEGKIIIRKIRGNNTDTRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 361  
Db 1147 NEGKIIIRKIRGNNTDTRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 1206  
Qy 362 LREQTKDINDNIIFQIOPMNNYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHN 421  
Db 1207 LREQTKDINDNIIFQIOPMNNYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHN 1266

Qy 422 YLVPVKOGNYASLLESTHGWGFPVSE 450  
Db 1267 YLVPVKOGNYASLLESTHGWGFPVSE 1295  
RESULT 11  
PCT-US97-15394-62  
; Sequence 62, Application PC/TUS9715394  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/15394  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US97-15394-62

Query Match 96.9%; Score 2317; DB 1; Length 462;  
Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 14 SLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLNFIFFDFKLGSS 73  
Db 26 ALLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLNFIFFDFKLGSS 85  
Qy 74 GEDRGKVIYQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVNNSGWSIGIIS 133  
Db 86 GEDRGKVIYQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVNNSGWSIGIIS 145  
Qy 134 NFLVFTLKQNEDESEISINFSYDISNAPGYNKFFVTVTNNMGMNMIYINGKLIDTIKV 193  
Db 146 NFLVFTLKQNEDESEISINFSYDISNAPGYNKFFVTVTNNMGMNMIYINGKLIDTIKV 205  
Qy 194 KELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYIFAKELDGKDINILFNSLOY 253  
Db 206 KELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYIFAKELDGKDINILFNSLOY 265  
Qy 254 TNVVDYWGNDLRYNKEYYVMVNDYLNRYMYANSRQIVFNTRRNNDFNEGKIIIRKR 313  
Db 266 TNVVDYWGNDLRYNKEYYVMVNDYLNRYMYANSRQIVFNTRRNNDFNEGKIIIRKR 325  
Qy 314 GNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIGLREQTKDINDNI 373  
Db 326 GNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385

QY 374 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHNLYLPTVKQGNYA 433  
DB 386 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHNLYLPTVKQGNYA 445  
QY 434 SLESTSTHWGFVPVSE 450  
DB 446 SLESTSTHWGFVPVSE 462

RESULT 12  
US-08-704-159-62  
; Sequence 62, Application US/08704159  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08704,159  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 703-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-704-159-62

Query Match 96.9%; Score 2317; DB 11; Length 462;  
Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 73  
DB 26 ALLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 85

QY 74 GEDRGKVIIVTQENIVYNSMYESFISFWIRNKWWSNLPCTTIIDSVKNSGWSIGIIS 133  
DB 86 GEDRGKVIIVTQENIVYNSMYESFISFWIRNKWWSNLPCTTIIDSVKNSGWSIGIIS 145

QY 134 NFLVFTLKQNEDESEGSINFSYDISNNAFCYKWFVTVTNMGMNKKIYINGKLIDTIKV 193  
DB 146 NFLVFTLKQNEDESEGSINFSYDISNNAFCYKWFVTVTNMGMNKKIYINGKLIDTIKV 205

QY 194 KLTGTFNFKTTTFEINKTPDTGLTSSDNNINMWIRDFYFAKELGDKDNIILFNSLQY 253  
DB 206 KLTGTFNFKTTTFEINKTPDTGLTSSDNNINMWIRDFYFAKELGDKDNIILFNSLQY 265

QY 254 TNNVKDYWGNDLRNKEYYVWVIDYLNRYMANSRQIVFNTRRNNDFNEGYKIIIKRIR 313  
DB 266 TNNVKDYWGNDLRNKEYYVWVIDYLNRYMANSRQIVFNTRRNNDFNEGYKIIIKRIR 325

QY 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKMETMYADNHSTEDIYAIGLREQTKDINDNI 373  
DB 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKMETMYADNHSTEDIYAIGLREQTKDINDNI 385  
QY 374 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHNLYLPTVKQGNYA 433  
DB 386 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRLLGGDWYRHNLYLPTVKQGNYA 445  
QY 434 SLESTSTHWGFVPVSE 450  
DB 446 SLESTSTHWGFVPVSE 462

RESULT 13  
US-10-271-012-62  
; Sequence 62, Application US/10271012  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/271,012  
; FILING DATE: 15-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 703-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:  
US-10-271-012-62

Query Match 96.9%; Score 2317; DB 26; Length 462;  
Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 73  
DB 26 ALLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 85

QY 74 GEDRGKVIIVTQENIVYNSMYESFISFWIRNKWWSNLPCTTIIDSVKNSGWSIGIIS 133  
DB 86 GEDRGKVIIVTQENIVYNSMYESFISFWIRNKWWSNLPCTTIIDSVKNSGWSIGIIS 145

QY 134 NFLVFTLKQNEDESEGSINFSYDISNNAFCYKWFVTVTNMGMNKKIYINGKLIDTIKV 193  
DB 146 NFLVFTLKQNEDESEGSINFSYDISNNAFCYKWFVTVTNMGMNKKIYINGKLIDTIKV 205

```

QY 194 KELTGINFSTKTIPEINKIPDTGLTSDSDNINMWIRDFYIFAKELDGKDINILFNSLOY 253
Db 206 KELTGINFSTKTIPEINKIPDTGLTSDSDNINMWIRDFYIFAKELDGKDINILFNSLOY 265
QY 254 TNVVDYWGNDLRYNKYYVNVNIDYLNRYMYANSQIVFNTNRNNDNFEGYKIIKRIR 313
Db 266 TNVVDYWGNDLRYNKYYVNVNIDYLNRYMYANSQIVFNTNRNNDNFEGYKIIKRIR 325
QY 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGREQTKDINDMI 373
Db 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGREQTKDINDMI 385
QY 374 IFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHHNLYVPTVKOGNTA 433
Db 386 IFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHHNLYVPTVKOGNTA 445
QY 434 SLESTSTHWGFVPVSE 450
Db 446 SLESTSTHWGFVPVSE 462

RESULT 14
US-09-791-537-59796
; Sequence 59796, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59796
; LENGTH: 1285
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-791-537-59796

Query Match 76.3%; Score 1824; DB 21; Length 1285;
Best Local Similarity 77.9%; Pred. No. 5e-159;
Matches 352; Conservative 36; Mismatches 56; Indels 8; Gaps 4;

QY 2 TIPNIFSYTNNLSLLKDIINEYFNNDINDSKLSQNKNTLMDTSGYNAEVRVEGVQLN 61
Db 839 TIPNIFSYTNNLSLLKDIINEYFNNDINDSKLSQNKNTLMDTSGYNAEVRVEGVQLN 898
QY 62 PIFPDFKLSSGDDRGKVIIVTONENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121
Db 899 PIFPDFKLSSGDDRGKVIIVTONENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 958
QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNNAAGYKNKWFVTVTNNMGNMKI 181
Db 959 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNNAAGYKNKWFVTVTNNMGNMKI 1018
QY 182 YINGKLIDTIKVKELTGINFSTKTIPEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIKVKELTGINFSTKTIPEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 1078
QY 242 KDINILFNSLOYTNVVDYWGNDLRYNKYYVNVNIDYLNRYMYANSQIVFNTNRNNDNF 301
Db 1079 KDINILFNSLOYTNVVDYWGNDLRYNKYYVNVNIDYLNRYMYANSQIVFNTNRNNDNF 1138
QY 302 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 360
Db 1139 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 1193
QY 361 GLREQTKD-INDNIIFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418
Db 1194 GALTQPMDEIRKYGSIITQPCNTFDYASQIFLSSNATNRLGLISGYSFPLGGDYWF 1253

```

```

QY 419 RHNVLVPTVKOGNYASLLESTSTHWGFVPVSE 450
Db 1254 RHNVLVPTVKOGNYASLLESTSTHWGFVPVSE 1285

RESULT 15
US-09-791-537-16595
; Sequence 16595, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16595
; LENGTH: 1285
; TYPE: PRT
; ORGANISM: Clostridium botulinum phase d
US-09-791-537-16595

Query Match 76.2%; Score 1823; DB 21; Length 1285;
Best Local Similarity 77.7%; Pred. No. 6.1e-159;
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

QY 2 TIPNIFSYTNNLSLLKDIINEYFNNDINDSKLSQNKNTLMDTSGYNAEVRVEGVQLN 61
Db 839 TIPNIFSYTNNLSLLKDIINEYFNNDINDSKLSQNKNTLMDTSGYNAEVRVEGVQLN 898
QY 62 PIFPDFKLSSGDDRGKVIIVTONENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121
Db 899 PIFPDFKLSSGDDRGKVIIVTONENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 958
QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNNAAGYKNKWFVTVTNNMGNMKI 181
Db 959 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNNAAGYKNKWFVTVTNNMGNMKI 1018
QY 182 YINGKLIDTIKVKELTGINFSTKTIPEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIKVKELTGINFSTKTIPEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 1078
QY 242 KDINILFNSLOYTNVVDYWGNDLRYNKYYVNVNIDYLNRYMYANSQIVFNTNRNNDNF 301
Db 1079 KDINILFNSLOYTNVVDYWGNDLRYNKYYVNVNIDYLNRYMYANSQIVFNTNRNNDNF 1138
QY 302 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 360
Db 1139 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 1193
QY 361 GLREQTKD-INDNIIFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418
Db 1194 GALTQPMDEIRKYGSIITQPCNTFDYASQIFLSSNATNRLGLISGYSFPLGGDYWF 1253
QY 419 RHNVLVPTVKOGNYASLLESTSTHWGFVPVSE 450
Db 1254 RHNVLVPTVKOGNYASLLESTSTHWGFVPVSE 1285

Search completed: November 7, 2002, 14:52:41
Job time : 294 secs

```

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:44:06 ; Search time 15 Seconds  
(without alignments)  
212.562 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391  
Sequence: 1 MTTPFNIFSYTNNLLKDI.....NYASLLESTSHWGFVPVSE 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 28998 seqs, 7085409 residues

Total number of hits satisfying chosen parameters: 28998

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/paa/PTCT\_NEW\_COMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pcp.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pcp.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pcp.\*
- 6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pcp.\*
- 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	5.1	912	6	US-10-092-411A-2993
2	121.5	5.1	1335	6	US-10-092-411A-3716
3	119.5	5.0	930	6	US-10-092-411A-5314
4	117	4.9	691	6	US-10-092-411A-4675
5	116.5	4.9	1151	6	US-10-092-411A-3242
6	108	4.5	976	1	PCT-US02-05068-98
7	107.5	4.5	897	5	US-09-721-456-189
8	107.5	4.5	1007	5	US-09-721-456-187
9	107.5	4.5	1031	5	US-09-721-456-110
10	106	4.4	1010	6	US-10-092-411A-5178
11	105	4.4	10182	6	US-10-092-411A-3159
12	104.5	4.4	497	6	US-10-092-411A-5114
13	104.5	4.4	695	6	US-10-092-411A-4341
14	102	4.3	3696	6	US-10-092-411A-5080
15	100	4.2	527	6	US-10-092-411A-4465
16	100	4.2	804	6	US-10-092-411A-5218
17	99.5	4.2	670	6	US-10-092-411A-2940
18	99.5	4.2	1041	6	US-10-145-087A-498
19	99.5	4.2	1041	6	US-10-143-031A-498
20	99.5	4.2	1041	6	US-10-145-092A-498
21	99.5	4.2	1041	6	US-10-162-522A-498
22	99.5	4.2	1041	6	US-10-165-038A-498
23	99.5	4.2	1041	6	US-10-165-353-498
24	99.5	4.2	1041	6	US-10-170-481A-498
25	99.5	4.2	1041	6	US-10-172-039A-498
26	99.5	4.2	1041	6	US-10-145-016A-498

27	99.5	4.2	1041	6	US-10-145-088A-498	Sequence 498, App
28	99.5	4.2	1041	6	US-10-145-129A-498	Sequence 498, App
29	99.5	4.2	1041	6	US-10-165-353A-498	Sequence 498, App
30	99.5	4.2	4544	6	US-10-085-198-22	Sequence 22, Appl
31	99	4.1	956	6	US-10-092-411A-4452	Sequence 4452, Ap
32	98.5	4.1	422	6	US-10-092-411A-5230	Sequence 5230, Ap
33	98.5	4.1	806	6	US-10-092-411A-4314	Sequence 4314, Ap
34	97.5	4.1	428	6	US-10-092-411A-2942	Sequence 2942, Ap
35	97	4.1	509	6	US-10-092-411A-2996	Sequence 2996, Ap
36	96.5	4.0	430	6	US-10-092-411A-4712	Sequence 4712, Ap
37	96	4.0	1162	6	US-10-092-411A-4008	Sequence 4008, Ap
38	95	4.0	331	6	US-10-092-411A-3626	Sequence 3626, Ap
39	95	4.0	1151	6	US-10-060-019-31	Sequence 31, Appl
40	94.5	4.0	366	6	US-10-092-411A-4412	Sequence 4412, Ap
41	93.5	3.9	840	5	US-09-721-456-190	Sequence 190, App
42	93.5	3.9	872	5	US-09-721-456-221	Sequence 221, App
43	93.5	3.9	1183	6	US-10-092-411A-3530	Sequence 3530, Ap
44	93	3.9	267	6	US-10-092-411A-3520	Sequence 3520, Ap
45	93	3.9	478	6	US-10-271-145-2	Sequence 2, Appl

## ALIGNMENTS

### RESULT 1

US-10-092-411A-2993  
; Sequence 2993, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lyng Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 2993  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-2993

Query Match	5.1%	Score 122;	DB 6;	Length 912;
Best Local Similarity	20.6%	Pred. No. 0.26;		
Matches	96;	Conservative	72;	Mismatches 154; Indels 144; Gaps 23;
QY	3	IPFNFTSTNN-----SLKDIINEYFNINDSKILSLONKNTLVDTSGYNA--E	51	
DB	374	VFFDQMPPTASLKGHNPNVYTLLOCI--EYKG--REVELVRKIQKNTLANKKVVTSKEE	429	
QY	52	VSEEDVOLNIFPDPFKLGSGEDRGKVIYQENIVY-----NSMYSESFSIFWIRINK	107	
DB	430	IFQYGVV-----NELIDYNNLYKRRPNREIHSFGDKYLYENE	470	
QY	108	WYSNLPGTITDSVKNNSGWIGIISNLFVTL-----KQED-----SEQSINF	152	
DB	471	QSI-----MSIISKILLSNESVEGYSNVEFWLNNETSLDCKREKKEILLRMFSNISKISM	526	
QY	153	SYDISNNAFGYNKFWFVTVNMMGNMKIYI---NGKLIDITKVKELTGINFSTITFEI	209	
DB	527	VY-----GAAGTCKSLINHICNFYDKDVIVTANTAVDNIKKR-----IKLSNIKTSTI	578	
QY	210	NKIPDTGLTSDSDMINWIRDFYIPAKELGDKDINILFNSLQYTNVYKDWGNDLYNK	269	
DB	579	SK-----FLYNDKEKYDLIID---EAGTVSNKDMNRILENKOPELL-----LIVGD	622	
QY	270	EYMYNIDVLYNMYVAN---SRQIVENTRRNNDFNEGYKIIKIRGNTNDTRVRGGDI	326	



Db 623 NVQIESIDFGNWFETAKDVLKSNII-----NELTDMY-----RTKNDL 661

Qy 327 LYFDMTINKKAYNLFMKNEMMYADNHS-----EDYIAGLRB- 364

Db 662 LYFKWSREKKSNI---NEIINKKYSTRLDESIFNEFNKDEILCLNDYDGIYGINNR 718

Qy 365 --QKNDINDNIIFOIQP-----MMNTYYASQIFKSNFNGENI 400

Db 719 LQANKKNDVINGVKYKVGDPILFNETKYSPILE-NNLKGSII 763

RESULT 2

US-10-092-411A-3716

; Sequence 3716, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092,411A

; CURRENT FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 3716

; LENGTH: 1335

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-3716

Query Match 5.1%; Score 121.5; DB 6; Length 1335;

Best Local Similarity 19.8%; Pred. No. 0.46;

Matches 97; Conservative 74; Mismatches 193; Indels 127; Gaps 23;

Qy 10 YTNSSLKDIINEYFNININDSKILSLQNRKNTLVDSGYNAEVSSEGDVQLNPIFFDPK 69

Db 900 YANRTEL---INKQRTGNNYTVLLQDGTNTPL-----GWNINDVTYQN-- 942

Qy 70 LGSSSEGRGKIVYQNEIVNYSVY-----ESFSIFWIRINKWSNLP 114

Db 943 IGKQTSIGKYSVFETNGLYSIANGTKNQQLAPNTLANQAFNASKAYVVGK-----D 996

Qy 115 YTIIDSVKNNGSWISGIIISNPLVFTLTKQNSDSEOSINFSYD-ISNAPGYNKWFFVTYTN 173

Db 997 LYLYGTVNNRTGW---IAAKDLI---QNSTDAQSTPYNTYFVINNSKSY---FYMDPTK 1046

Qy 174 NMGMNMTIYINGKLIDITKYKELGINF-----SKTITEINKIPDGLIT 219

Db 1047 ANRSLKPYE-QFTYIKQKNGVWYVQQLDGGYVWIKSTDIVKIKYAYIGMTL 1105

Qy 220 SDDSNIN-----MWIRDFYIFAKELDGKIDINILFN--SLOYTNVVKDYGN 263

Db 1106 NNAINIQLSKYKQVQNEPLKWSNANSYQIK--NAMDTRKLANDSLKY-----Q 1154

Qy 264 DLRYNKEYVMYIDILNRYM---YANSRQIVNTNRNNDNFEGYKIIKIR-ONTND 318

Db 1155 FURLDQOYL-SQAALNKLKGGVLENGQGAFAAARKYGLNEIYLISHALVETGNGTS 1213

Qy 319 TVRGGDIL--YFDMTINNKAYNLFMKNEMMYADNHSDEIYAGLRQTKDINDNIIFQ 376

Db 1214 QLAGGDVSKGKFTTKTKGHYHVGIGA---PDNNALVGIKAYKNAGWTSVSKAIIIG 1270

Qy 377 QPMNTYYASQ--IFKSNFNGENISGICSTYRFLCGDVRHN----- 421

Db 1271 AKFIGNSVKAGQNTLYKMRWNPN-----PGTHQYATDINWANNVAQYKLFQYDKIGE 1324

Qy 422 ----YLVPTVK 428

Db 1325 VKGYFEITYK 1335

RESULT 3

US-10-092-411A-5314

; Sequence 5314, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092,411A

; CURRENT FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 5314

; LENGTH: 930

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-5314

Query Match 5.0%; Score 119.5; DB 6; Length 930;

Best Local Similarity 21.0%; Pred. No. 0.39;

Matches 98; Conservative 63; Mismatches 184; Indels 121; Gaps 23;

Qy 9 SYTN-----NSLLKDIINEYFNININDSKILSLQ-NRKNTLVDSGYNAEVSSEGDVQ 59

Db 224 SYKNIDEKISNQDELNLPINEVENKVRPLSTTSAPSSKRVTV-----NQLAAEGGSNV 278

Qy 60 LNPIFFDFXKLGSGEDRGKIVYQN-ENIVNYSVYSFISFWIRINKWSNLPYTTI 118

Db 279 NHLIKVTDQSITTEGYDSDGIKAHAENLIYDVTFE-----VD 317

Qy 119 DSVKNNGSWISGIIISN-----FLVFTLKQNSDSEOSINFSYDISNAPGYNKWFFV 169

Db 318 DKVSGDTMTVIDKNTVPSDLTDSFAIPKIKDN-SGEIATGYDNTINKQTYFTFDIV 376

Qy 170 TVTNMNMGNMKI--YI-----NGKLIDITKYKELTGINFSTITEINKIPD--TGLI 218

Db 377 DKYENIRAKHLKTSYIDKSKVPNNNTKLDVEYKLTSSVN--KITVEYQKPNENRTANL 434

Qy 219 TSDSDNIN-----MWIRDFYIFAKELDGKIDINILFNLSLOYTNVVKDYGNLRYNK 269

Db 435 QSMFTNIDTKNHTVEQTYIYNPLYSKAKE---TNVAISGNGDEGSTIIDD--STIKVYK 489

Qy 270 EYVMVNTDYLNR-YMVANSRQIVNTNRNNDNFEGYKIIKIRIGNTNDRVRGGDI-- 326

Db 490 VGDQNLPDSNRIVDYSEYDV-----TNDYAQ-----LGNNDVNINFGNIDS 534

Qy 327 LYFDMTI-----NKNAYNLFMKNEMMYAD-NHSTEDIYAGLRQTKDINDNIIFQI-- 378

Db 535 PYIIKVISKYPDKNDYDTTQQQVMTQTTINEYTGFE-----RTASYDNTIAPSTSSG 587

Qy 379 -----PMNNTYYASQIFK-----SNFNGENISGICSTGY 409

Db 588 QCGDLPEPKYKIGDYWEDVDKDGQIONTNDNEKPLSNVLVTLY 633

RESULT 4

US-10-092-411A-4675

; Sequence 4675, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092,411A

; CURRENT FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 4675  
; LENGTH: 691  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-4675

Query Match 4.9%; Score 117; DB 6; Length 691;  
Best Local Similarity 19.3%; Pred. No. 0.4;  
Matches 116; Conservative 84; Mismatches 176; Indels 224; Gaps 33;

QY 1 MTPFNIFSYTNNS-----LLKDIINEYFNINDSKIL----- 33  
DB 5 MSIPNL--PTNSTMINELCTLOSRTINIKGEVLITELIYDDYFFKNDWEHITAFNKQF 62  
QY 34 --SLQN-----VDTSGYNAE-----VSEE-----GVOLNPIPF 67  
DB 63 QDSIKYRDKRKNVFRFKSKNLNLEFKYLFKLIVKEDWSLNFNTGAVKLNKIAF- 121  
QY 68 FKLGSGGDRGVIVTQENIIVYNSMYSESISFWIRN-----KWV--SNLP-----G 114  
DB 122 -----FNEVYNLSLLDCLDINTLEKWFNWLPIENNPIKERS 159  
QY 115 YTIIDSVKNNSGSIGIISNLFVFTLKONEDSEQISNFSYDISN--NAPGYNKWFVTV 171  
DB 160 STIVFGDYKSLASFLKKNYINLIKIDKREWEKDKWDIRNLEKYGLSYNK-----TL 215  
QY 172 TNNMGNMKIYKINGLIDITKVKELTGINFESKTITFEINKIPDGLITSDSDNNMIRD 231  
DB 216 TGN-----YLFKEIESIKRELA-----KKYKLNLTGCD--TAFATAR 253  
QY 232 FYI-----FAKELDKIDINLFNSLQ-----YTNVVKDYWGN--DLRYKEYYMWNI 278  
DB 254 FYIRVLTREFONISKKNKERNLSNELDRCHIEAYTEFLFEYAANKHLQSTKNFVREELKT 313  
QY 279 LNYM-----YANSQIVF-----NTRNNNDNFEGYKIIIKIRGNTD-- 318  
DB 314 IRRFLNDIITQNTVAIAPYQDIRFLYIPQDLPKHEKNSQIDYIPDFLEQOLFHEINDLH 373  
QY 319 -----TRVRGGDILYFDMTINNK-----YNL-----FMKNETMYADNH 352  
DB 374 KDLIPVWIAFKTGLRISDVL-----TLQNNCLAKYNGKYSIIITDIATKTVKGRIPIDN- 428  
QY 353 STEDIVAIGLRE-----QTKDIN--DNIIFOI--OPMNTYYIASQIFKSNF-----GENI-- 400  
DB 429 KLADITAVLIADSKSKSTKDNPNPNYIFAIRYKGRKGMPFTQHMVRAHLNHLSTKTNIID 488  
QY 401 --SGICSIGYRFLGSDWRHNYLPTVKQG-----NYASLLEST 439  
DB 489 EQGEIHFKTHQF-----RHTYAVKLLNGGADILITQELLAHSSPEWTLRYAKLLDDT 541

RESULT 5  
US-10-092-411A-3242  
; Sequence 3242, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 3242  
; LENGTH: 1151  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-3242

Query Match 4.9%; Score 116.5; DB 6; Length 1151;  
Best Local Similarity 17.9%; Pred. No. 0.81;  
Matches 97; Conservative 90; Mismatches 175; Indels 181; Gaps 27;

QY 3 IPFNIFSYTNNSLLKDIINEYFNINDSKILSLQNRKNTLV-----DTSGYNAEYSEGDV 58  
DB 72 LPSSVPPTSTNFAIVSAKE-----DEIIANTLQQQYTKLKIYNDVKOMRQNVDES 125  
QY 59 QUNPIPFDP--FKLGSGGDRGV-----IVTQN-----ENIVYNSMYSESIS 100  
DB 126 EIN--FPSKNFNLGTFQDTFGVDSNVATHOSQEQPMYTSNLLFYTVYDNYHVSALNFK 183  
QY 101 FWIRINKWVSNLPGYTIIDSV-----KNNSGWSIGI----- 131  
DB 184 FMKRLNE--VGPIIFVINQDKHNEBELITETFKSRVEKSIDKDWIKLODTYVVSKEFDH 241  
QY 132 -----ISNLFVFTLKONEDSEQISNFSYDISNAPGYNKWFVTVTNNMGNMKIYIN 184  
DB 242 PONEIDKLSNLFVMDQHRESTEDYVNETIQFITDAQ-----YIYIONEMO----- 287  
QY 185 GLIITIKVKELTGINFESKT--ITFEINKIPDGLITSDSDNNMIRDIFYIFAKELDKGD 243  
DB 288 -SLDPTLQINE--EQFEAYIQFOQNO-----EVSARAQLLN-----D 322  
QY 244 INILFNSLOYTVNVKDYWGN-----DLRYKEYYMWNI--DY--LNRYMYANSQIVTENT 294  
DB 323 SNOLFNYLKQKR--KDILDNAYIMTYDMRESLRNYLESMTATDFKVNCF-----FNK 371  
QY 295 RNNNDNFEGYKIIIKIRGNTDNRVRGG-----DILYEDMTINNKAYNLFMKNE 345  
DB 372 RKXKEEQ-----IKRLNEATQLOEQVNOQVROPLREDSFTRFNKAVNEKILNQ 425  
QY 346 -----TWADNHSTEDIYAIGLREQT--KDINDNIIFQIOPMNTYYIASQIFKSN 394  
DB 426 EYDVPFSLISELYQQTSTISNTYVLTFSDEVIKALNKXIENESTP-----LFEEA 475  
QY 395 FNGENISGICS-----IGYRFLGSDWRHNYLPTVKQGNYSALLSSTS 440  
DB 476 VNHQVYNELSSDENEDRYEYDRIELNTLKDSLTSHTNKHYYI----HLDDSLDKLIGTE 532  
QY 441 THW 443  
DB 533 THF 535

RESULT 6  
PCT-US02-05068-98  
; Sequence 98, Application PC/TUS0205068  
; GENERAL INFORMATION:  
; APPLICANT: Diversa Corporation  
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY  
; FILE REFERENCE: 09010-108W01  
; CURRENT APPLICATION NUMBER: PCT/US02/05068  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: US 60/270,495  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: US 60/270,496  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: US 60/291,122  
; PRIOR FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 321  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 88  
; LENGTH: 976  
; TYPE: PRT

ORGANISM: Bacterial  
PCT-US02-05068-98

Query Match 4.5%; Score 108; DB 1; Length 976;  
Best Local Similarity 21.8%; Pred. No. 2.4;  
Matches 101; Conservative 70; Mismatches 157; Indels 136; Gaps 27;

QY 2 TIPNI-FSTNNSLLKDI--NEVENINDSKILSLQNRKNTLVDTSGYNAEYSEGD 57  
DB 206 TVKEVDLKDASMLPEKHLNESEIEVLPKATSL--TNVVDPT-----EVNEP-- 256  
QY 58 VQLNPFP-----FDFKLGSSGDRGVIVTQNEINIVNSMYESFSFWI-----RIN 106  
DB 257 -KLWVFPQIELFPDVTNLGSE-----INELEGNLSMNWDYSDPNABFYVYKLEBQ 310  
QY 107 KWSNLPD-----YTIIDSVKN--SGWSIGIISN---FLVETLQN-----EDSE 147  
DB 311 EYLFEVGKTRKSYTIENTKQEFKFSIAINVTANKESGLVVLKKNKLDLDES 370  
QY 148 QSINFSYDISNAPGVNKFV--TVTNMMGMNKIYINGKLIDTIKYLKELGINSFKTI 205  
DB 371 DSISATYVDTNELKD-WNYTSSVTFEV---LKKGINSVEYII--SQTQNSFSTEF 424  
QY 206 T-----FEINKIPDTGLTSDSDINMKWIRDYIFAKELDGKINI-----LFNSIQYN 255  
DB 425 TGRFVLEKIAIRVANGFESKIN-----EISRDITITSLNPLTSSMTMYL 473  
QY 256 VKDYWQND-----KNETMYADNHTEDIYAIGREQT-KDIN-DN 372  
DB 474 FIRSYFDTDGGVDFSGVAEKVDYLSGLGVDVWVLPFNPKSKSYHGVDYDVEPDY 533  
QY 279 LNRVYANSQIVENTRRNNDNEGYKIIKIRGTNTDTRVGRGDILYFDMTINNAY 338  
DB 534 GTLQDLNMIKVL-----NEN-----GIKVMYDLVNVHTSDTHPWFLDAY--ENTTNSPY 582  
QY 339 NLFM-----KNETMYADNHTEDIYAIGREQT-KDIN-DN 372  
DB 583 NYIIMSLDEQNKNNHMYKYNKSGQIVWVYFGLFDSMPLNDN 626

RESULT 7  
US-09-721-456-189

Sequence 189, Application US/09721456  
GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin B.  
Andrews, William H.  
TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/721,456  
FILING DATE: 22-Nov-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 897 amino acids  
TYPE: amino acid  
STRANDEDNESS: <unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 189:  
US-09-721-456-189

Query Match 4.5%; Score 107.5; DB 5; Length 897;  
Best Local Similarity 18.5%; Pred. No. 2.3;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;

QY 7 IFSY-TNNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGY----- 48  
DB 217 LFSYTDNKCVTQFINEFFYNILPKDFLTGRNKNFKQKKVYKVELNKEELHKLLEK 276  
QY 49 --NAEVS-EGDVLNPPIPPDFK-----LGSSGDRGVIVTQNEINIVY-NSMYES 96  
DB 277 INREISWQVETSAAKHFYFDHENIYVLKLLRWTFD---LVVSLRCFFYVTEQKS 333  
QY 97 FSIWFIRINKW-----VSNLPGYTIIDSVKNNSGWSIGIISNPLVFTLKQNEDEQS 149  
DB 334 YSKTYVYKKNIDVIMKMSIADLKKETLAEVQE-----KEVEENKKS 375  
QY 150 INFSYDISNAPG-----YKNWFFVTVTNMMGMNKIYINGKLIDT-IK 192  
DB 376 LGP-----APKRLIPKKTTFRPFIMFNK-----KIVNSDKTKTLTNTKLLNSHLM 424  
QY 193 VKELTGINFSTITFEINKIPD-----TGLITSDNIN--- 226  
DB 425 LKTLKRNMFDPFGFAVFNDDVMKKYEEFVCKWQVGPQLFFATMDIEKCYDSVNRK 484  
QY 227 -----MWIRDFYIFAKELDGKIDINILFNSLOY-TNVVKDYNGNDLRYNKYYWNI 276  
DB 485 LSTFLKTKLLSSDFWIMTAQILKRNKNIVIDSKPRKEMKDY-----FKQFKQIAL 538  
QY 277 DYLNRYMYANSRQIVENTRRN--NDFNEGYKIIKIRGTNTDTRVGRGDILYF----- 329  
DB 539 E-----GGGYPTLFSVLENEQNDLNAKTLIVEAKORN-----YFKKDNLL 579  
QY 330 DMTINNKAYNLFKNKNTMTADNHTEDIYAIGREQTKIDNINIFQTPMNTYYASQ 389  
DB 580 QPVINICQYNYINFNGKFY-----KOTKGIPOGLC--VSSILUSSFYIATL 622  
QY 390 -----IFKSNFNGENISGICSGTGYRFRFLGDDWRHNYLVTVKQGNYSALLF---ST 439

Db 623 EESSIGFLRDESMNPEN-----VNLMLRLTDD-----YLLITTOENNAVLFIKLIYN 672

Qy 440 STHWGF 445

Db 673 SRENGF 678

RESULT 8

US-09-721-456-187

Sequence 187, Application US/09721456

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin B.  
Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/721.456  
FILING DATE: 22-Nov-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/974.549A  
FILING DATE: 19-Nov-1997  
APPLICATION NUMBER: US 08/724.643  
FILING DATE: 01-Oct-1996  
APPLICATION NUMBER: US 08/844.419  
FILING DATE: 18-Apr-1997  
APPLICATION NUMBER: US 08/846.017  
FILING DATE: 25-Apr-1997  
APPLICATION NUMBER: US 08/851.843  
FILING DATE: 06-May-1997  
APPLICATION NUMBER: US 08/854.050  
FILING DATE: 09-May-1997  
APPLICATION NUMBER: US 08/911.312  
FILING DATE: 14-Aug-1997  
APPLICATION NUMBER: US 08/912.951  
FILING DATE: 14-Aug-1997  
APPLICATION NUMBER: US 08/915.503  
FILING DATE: 14-Aug-1997  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-Oct-1997  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-Oct-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 187:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1007 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 187:  
US-09-721-456-187

Query Match 4.5%; Score 107.5; DB 5; Length 1007;  
Best Local Similarity 18.5%; Pred. No. 2.7;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;

Qy 7 IFSY-TNNSLLKDIINEYFNNINDSKILSLONKNTLVDTSGV----- 48  
Db 347 LFSYTTDNKCVTQFINEFFYNILPKDELTOGRNRKNCKKKYKVEINKHELHKNLLLEK 406  
Qy 49 --NAEVS-EEGDVQLNPIFFDFK-----LGSSGEDRGKVIIVTQENIVY-NSMYS 96  
Db 407 INTREISWMOVETSAKHFFYFDHENIYVLKLLRWIFED---LVWSLIRCFYVTEQKS 463  
Qy 97 FSIWFIRINKW-----VSNLPGYTIIDSVKNNSGWSGIISNLVLTQKNESEOS 149  
Db 464 YSKTYYYRKNIVDMKMSIADLKKTAEVQE-----REVEEWKS 505  
Qy 150 INFSYDISNAPG-----YNKWFFVTVTNMMGNMKIYINGKLIDT-IK 192  
Db 506 LGF-----APGKRLIPKKTFRPIMTENK---KIVNSDRKTTKLTNTKLLNSHLM 554  
Qy 193 VKELTGINFSTITFEINKIPD-----TGLTSDSDNIN--- 226  
Db 555 LKTLKNRMFKDPFGFAVENYDDVMKKYEEFVKWQVQPKLFFATMDIEKCYDSVNRK 614  
Qy 227 -----MWIRDFYIFAKELDGKDNILFNSLOY-TNVVKDYWGNDLRYNKEYTYMVI 276  
Db 615 LSTPLKTKLLSSDFWMTAQILARKNNIVDSKNFPKKEMKDY-----FRKFQKIAL 668  
Qy 277 DYLNRYMYANSRQIVFNTRN--NNDNEGYKIIKIRGNTNTRVRGGDILIF----- 329  
Db 669 E-----GGQYPTLFSVLENEQNDLNAKKTLLIVEAKQRN-----YFKKDNL 709  
Qy 330 DMTINKAYNLFMKNETMYADNHSTEDIYALGLREQPKDNDNIIFQIOPMNTYTYASQ 389  
Db 710 QPVINICQYINENGKPY-----KQYKG:POGLC--VSSILSSFYAIL 752  
Qy 390 -----IFKSNFNGENISIGSICTYRFRGLGSDWYRHNLYLVPTVKQGNYSALLE---ST 439  
Db 753 EESSIGFLRDESMNPEN-----VNLMLRLTDD-----YLLITTOENNAVLFIKLIYN 802

Qy 440 STHWGF 445

Db 803 SRENGF 808

RESULT 9

US-09-721-456-110

Sequence 110, Application US/09721456

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin B.  
Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/721,456

FILING DATE: 22-NOV-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997

APPLICATION NUMBER: US/08/724,643

FILING DATE: 01-OCT-1996

APPLICATION NUMBER: US/08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US/08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US/08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US/08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US/08/911,312

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US/08/912,951

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US/08/915,503

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

APPLICATION NUMBER: WO PCT/US97/17885

FILING DATE: 01-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph Ted

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002610US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 110:

SEQUENCE CHARACTERISTICS:

LENGTH: 1031 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 110:

US-09-721-456-110

Query Match 4.5%; Score 107.5; DB 5; Length 1031;  
Best Local Similarity 18.5%; Pred. No. 2.8;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;

QY 7 IPST--TNNLLKDIINEYFNIN--SKILSLQNRKNTLVDTSGY-----48  
DB 348 LFSYTTDKKCVTQFINFPYILPKDLTGRNKNFQKKYKVELNKHHLHKNLLEK 407  
QY 49 --NAEVS--EGDVQLNPFPDFK-----LGSSGDRGKVIIVTQNEIVY--NSMYES 96  
DB 408 INTRISWQVETSAKHYYFDHENIYVLKLLWIFED--LVVSLIRCFYFVTEQOKS 464  
QY 97 FSISEWIRNKW-----VSNLPGYTIIDSVKNNSGWSIGIISNLFVTLKQNESEQS 149  
DB 465 YSKTYVYRKNIWDVIMKMSIADLAKETLAEVQE-----KEVEEWKKS 506  
QY 150 INFSDYISNAPG-----YKWFVFTVNTNMGMNKIYINGKLIDT-IK 192  
DB 507 LGF-----APGKURLPKKTTTPRINTFNK-----KIVNSDRKTKLTNTKLLNSHLM 555  
QY 193 VKELIGINFSTKITFEINKIPD-----TGLITSDSDNTN---226  
DB 556 LKTLNRNFKDPFGFAVENDYDMVKYBEFVCKWKVQGPQKLPFATMDIEKCYDSVNRK 615  
QY 227 -----MWLRDYIIFAKELDKDINILNSLOY-TNVVKDYWGNDLYNKEYYMWNI 276  
DB 616 LSTFLKTKLSSDFWIMTAQILKKNKNIIVDSKNFRKKEMKDY-----FRQFKQIAL 669  
QY 277 DYLNRYMYANSQIVFNTFRN--NDFNEGYKIIIRKIRGNTND\*RVRGGDILYF-----329

DB 670 E-----GGQYPTLFSVLENEQDNLNAKKTILVEAKQRN-----YFKKDLML 710  
QY 330 DMTINKAYNLFMKNETMYADNHSTEDIYAIGLREQTQKIDNINIIFQIOPMNTYVYASQ 389  
DB 711 QPVINICQYNYINFNGKFY-----KOTKIGIPOGLC--VSSILSSFYATL 753  
QY 390 -----IFKSNFNGENISGICISGTYRPLGGDWYRHHYLVPTVKQGNVYASLLE---ST 439  
DB 754 EESSLGFPLRDESNPNPN-----VNLMLRLTDD-----YLLITTCENNAVLFIKLIINV 803  
QY 440 STHWGF 445  
DB 804 SRENGF 809  
RESULT 10  
US-10-092-411A-5178  
; Sequence 5178, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 5178  
; LENGTH: 1010  
; TYPE: PRT  
; ORGANISM: staphylococcus epidermidis  
US-10-092-411A-5178

Query Match 4.4%; Score 106; DB 6; Length 1010;  
Best Local Similarity 19.7%; Pred. No. 3.4;  
Matches 75; Conservative 60; Mismatches 100; Indels 146; Gaps 19;

QY 8 PSYNNLLKDIINEYFNNI-----NDSKILSLQNRKNTLVDTSGYNAEYSEGDVQLNPI 63  
DB 674 YEYENK-TKDKLNLNKKLKIEMNDQKHLT-ENLTQTSKEIN--NLELMEKEMQ----725  
QY 64 PFDFKLGSSGDRGKV---IVTQNEI-----VYNSMYESFISIFWIRINKWVSNLPGY 115  
DB 726 -----QLGFESYDQVKSADLSAQKDEIEREINIYKNKYQSYE---IEINRLKELVKCK 776  
QY 116 TIID-----SVKNNSGWSIGIISNLFVTLKQNESEQSINFSDYISNAPGYNKWFFV 169  
DB 777 KLLNLEELRQSIKENT-----LKLDETNSQIATISYKIDNNSKNFK-----818  
QY 170 TVTNMGMNKIYINGKLIDTPIKVKELTGINFSTKITFEINKIPDTGLITSDSDNINMWI 229  
DB 819 -IKN-----LIQLDDELKQV-----833  
QY 230 RDFYIFAKELDKDINILNSLOYTNVVKDYWGNDLYNKEYYMWNIIDYLNRYMYANSRQ 289  
DB 834 KEIFLSSLEILAGK-----NDYKLTLENYVL-IYLEKIFIQANCR 872  
QY 290 IVFNTRRNNDNFNGYKIIIRKIRGNTND\*RVRGGDILYFDMIINKAYNLFMKNETMYA 349  
DB 873 LSFMS-----GNRYQL-----IRRETISLGLSGLEIDVFD-----FHSNKRHI 911  
QY 350 ONHSTEDIY-----AIGLRE 364  
DB 912 SLSLGGEFQFASLALGLSE 932

RESULT 11

```
US-10-092-411A-3159
; Sequence 3159, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-3159

Query Match      4.4%; Score 105; DB 6; Length 10182;
Best Local Similarity 21.2%; Pred. No. 71;
Matches 121; Conservative 71; Mismatches 202; Indels 176; Gaps 28;

QY 6 NIPSYNNLLKDIINEYNNNDKSLQNRKN--TLVDTSGVNAEVSSEGDVQLNP- 62
D 1313 NVSSVTVTRKGDLEKFNSSNNNNFLRWNRDYYTISTSSKIQRNVDGGLENSPD 1372
QY 63 -----IPFPDFKLGSGEDGRGVIVTQ-----NENIVYN-----SMYEFSSFSFWIRI--- 105
D 1373 HKSLIYKYDGLTRQINTNDVLTLLQATAKNSLNRSNINSNEKQALAEKSGNGYSKSIIRD 1432
QY 106 ---NKVSNLPGYTIIDSVKNSGWSIGIISNLFVTLKQEDSEDSQSFNSFYDLSNNAPG 162
D 1433 DGEKSYLLNSFTQVLDLVEPDNGGGRGVSHSNVT---YNEKSSIVNGQVPPANGASA 1489
QY 163 YN--KWFVTVTN--MMG-----NMKIYINGKLIDIRY 193
D 1490 FNIDKVVKANAANGIMGVIYKAQLAPYSPKGYIEKLQNLSTNNVINVYFVPSDKV 1549
QY 194 KE-----LTGINFSTIIFTEINKIPDTGLTISDSNINNMWTRDYIFAKELDG 241
D 1550 NPSITVGNVDHHTVYSGETFKNTINVNDYGLNVTASTSDS-AITMTNN-----NELVG 1603
QY 242 KDINILFNSLQTVNYK--DYWGNDLRYNKEYIMVNIIDYLN---RYMYANSRQI---VFN 293
D 1604 QAPNVT-NSINKTVKATDKSGNESIYS---FTYNIKPLNEKYRIITSSNQTPVRISN 1659
QY 294 TRNNN-----DEN-----EGYKIIKIRGNTNTRVGGDILYFDMTINKAYNLFMK 343
D 1660 IQNNANLSIEDQNRVKSSLSMTKILGTRNYVNESNDVRSQVYKVNKSGNNATVNV--- 1716
QY 344 NEMWAD-----NHSSTEDIYA----- 359
D 1717 -TTTFSDGTNTITVPVKHVLLEVPTTTRTVGGQFTFGKTSNDFESLRTGGPYDAR 1775
QY 360 -IGLREQTKDINDN-----IIF-QIQPM-NNTYYAS-----QIFKSNFNGEN 399
D 1776 IYVWNNQGPDIINSQIGRDLTLHAEIFFGGETTPRKDITVYKLSQSIKQIYETTING-- 1833
QY 400 ISGICSGIYRFLGGDWYRHNVLYPTVKQ 429
D 1834 -----RFNSSGDAIPGNF-VQAVNQ 1852

RESULT 12
US-10-092-411A-5114
; Sequence 5114, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
```

```
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 5114
; LENGTH: 497
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-5114

Query Match      4.4%; Score 104.5; DB 6; Length 497;
Best Local Similarity 21.3%; Pred. No. 1.7;
Matches 83; Conservative 51; Mismatches 153; Indels 103; Gaps 20;

QY 22 EYFNINDSKILSLQNRKNTLVDTSGVNAEVSSEGDVQLNPFFDFKLGSGEDGRKVI 81
D 129 EGFNLYHFNDLTQLDKH-----DVSGYK-----FNLESIASINFTSGTGPQKAVPQ 175
QY 82 VQQENIVNYSMEYSFISFWIRINKWSNLPGYTIIDSVKNSGWSI---GIISNLFVF 138
D 176 TFNNHLASAKGCKQSIGFE---QNTVWLSVLPYHI-----SGLSVILRAVIEGTVR 225
QY 139 TLK--QNEDESGSINFSYDISNNA--PGYNKWFVTVTNMMGNMKYING-----KLID 189
D 226 LVKKTQTDMLTQIK-TYPITHMSLVPTQTLKWLMDAGLTQPSLEKILLGAKLSPOLIE 284
QY 190 ---TIKVKELTGINFSKTITFEINKIP-----DTGLITSDSNINNMWIRDFYIFAK-E 238
D 285 QALTVELPYNSFGWTEFCQSFLTASPCMLKERPT--VGKPSENVKIKNPAYGHGE 342
QY 239 LQGDINILFNSLQTVNYVVDYWGNDLRYNKEYIMVNIIDYLNRYMYANSRQIVENTRNN 298
D 343 LLIKGENVM-NGYLYPKYLKDTFND-GYFQTGDAEID----- 379
QY 299 NDFNEGYKIIKIRGNTNTRVGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIY 358
D 380 ---DEGYVIYDR---RKDLISGGENIY-----PYQI-----ETIAKDFEGIEDAV 420
QY 359 AIGLREQT-----KDINDNIIFQ 376
D 421 CVGISDDTWGQVPILYVYTNQDINOTELIE 450

RESULT 13
US-10-092-411A-4341
; Sequence 4341, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 4341
; LENGTH: 695
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-4341
```

```

Query Match
Best local similarity 4.4%; Score 104.5; DB 6; Length 695;
Matches 100; Conservative 60; Mismatches 138; Indels 161; Gaps 25;

QY 10 YTNNSLLKDINEY-----FNNINDSKIL-SIQNRKNTLVDTSYNAEVSSEGVDQLN 61
Db 185 YGDNVLKDIAPYEAIRFLTKDMIKDSKIIFKDLNLYKSHFDQRKYRQLES----- 237
QY 62 PIFPFDKLGSGGDRGKAVITQENIYNSMYSPSISFWIRINKWWSNLPGYTIID-- 119
Db 238 -----TFKINTGQE-----FLNVTETSSQHVKKIHRHLGSEKOIDIA 279
QY 120 SYKNSGWSIGIISFLVF-----TLKQNEIDSEQSI-----NFSY 154
Db 280 SRLKESRWMLIRKNKYLFCNCKTYEYKRTMSYIKNLQIPVDDDSIFETSPHGKNFSG 339
QY 155 DISNNAPGV-----NKWFFVVTNNM-----GNMK----- 180
Db 340 D-----PKIALAIRQYDHHKLIIVSTNSLVDMEIRKRYGTFPVRFGSEKIKTFPRCKY 394
QY 181 IYING-----KLIDTIKVKELTGINFSKITITFEINK-----IPDTG-----LI 218
Db 395 VFINGMSDKVYKSSDQIFVQVTHGFPLKKWVN-DLNEQHEROQOLEAFIPRMKKWDYIL 453
QY 219 TSDSDINNMWRDFVIFAKELDKGKINILFNSLOYTNVVKDYW--GNDLRNKEY---YM 273
Db 454 TSDINTTLESAPML-----NKNPNL--KVLEYGAPKNEYLINNNLOEROLOLKYM 505
QY 274 VNIDYLNRYM-----ANSRQIVENTRRNND---FNEGYSKIIIK-----RIRGN 315
Db 506 YKIDDDKKYILCYPTWRGNORKEV--TQINLKDLLKYLPEYIIVKLHPNESHLTRYN 563
QY 316 TNDTRYRGGDILYFDMTINNKAYNLFMKNETMYADNHST 354
Db 564 QIDNRH----CYFNELV--IQELYILSECMTIDYSST 596

RESULT 14
US-10-092-411A-5080
; Sequence 5080, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092.411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-5080

Query Match
Best local similarity 4.3%; Score 102; DB 6; Length 3696;
Matches 92; Conservative 69; Mismatches 166; Indels 148; Gaps 21;

QY 6 NIFSNTNN---SLLKDII--NEYFNINDSKITLSQN-----RKNTLVDTSYNAEVS 53
Db 238 NGFTWNGKGVLLNSVLERTSVFSDAPKNYQADNVVALGRKNDPNDDHGFN---G 294
QY 54 EEGGVQLNP-----IPFPDFKLGSSGDRGKAVITQENY-----IYNSMYESFSI-- 99
Db 295 IEKEFSVNPNSEIIFSNFTMTAKNRKGTQVILRNAENNQBIASDTQGGGVYFLFKLPD 354
QY 100 -----SFWIRINKWWSNLPGYTIIDSYKNSGWSIGIISFLVFTLKQN 143

```

```

Db      355  NVNRKLVQFLPMNEIHSDFKRIQQLHDGTYRYFFIDTIGVNSGSHLYVSRQNVKVKNG 414
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      144  EDSEQSI-----NFSYDISNNAPGYNNKWFVTVTNMMGMKIYVINGKLIDTIKYKELT 197
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      415  KEPEVNTRIENNGNFAAIIQGNELTYK---VTLPENFE-----YVD-----NSTEVSFVN 461
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      198  GINFSKIT-PEINKIPDTCGLITSDSDNINMWIRDFVIFAKELDGKDNILFNSLQVTNV 256
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      462  GNPVNSTVNPFSVNFDEQNHILFFSSGNLNG-RSAQDVAREFLPNKILNI----- 510
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      257  VKDYWGNDLRYNKEYYKVVNTDYLNRYMYANSROIVNTRNNNDNFNEGKYLIIKIRGNT 316
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      511  -----RY--KLREVNT-----STPREVTFNEAIKYKTFSEYI-----NT 543
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      317  NDRVRGGDILYFDMTINNK-----AYNLFMK-----N 344
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      544  NDNTVTCQQPFSINVIMNKDDLSEQVKNDIIPSNYTLASYNKYNKLKERAQTVLDEETN 603
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      345  ETMYADNHSTEDIYALGRQET-----KDINDNIIIFIQPMNNTYYASQI 390
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      604  NTPFNQRYSTQIDDLHLHQLTTLINEVSASREIND---KAOEMTDAVYDSTEL 654
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 15
US-10-092-411A-4465
; Sequence 4465, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 4465
; LENGTH: 527
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-4465

Query Match 4.2%; Score 100; DB 6; Length 527;
Best Local Similarity 20.1%; Pred. No. 3.7;
Matches 93; Conservative 73; Mismatches 187; Indels 110; Gaps 22;

```

Qy	4	PNFTSFYTNNSLAKDI---INEFNFINDSKILSLONRKNLTVDTSYNAEV---SEBGD	57
Db	117	PFQOAITGONTFSKIHFSQGYLMWVEDYKSTIQRR--FVEDDRGFISAVRTYTPDGD	174
Qy	58	VQLNPPIFPDFKLGSS--GEDRGKVIIVQEN-----IVYNSYESFSISFWIIRINK	108
Db	175	NNKHYFSKD---GEEIFVEDLNVTYINKNFQSKFRVTVYSMAELIEEFQSYVERE	231
Qy	109	VSNLPGVTTIDSVKNNCGWSIGTISNPLVFTL--KONEDSEQINSFYDISNAPGYNK	166
Db	232	LNEDDSVIVASDERHSMARTIDASSLCSFSIFERNKVVYQDL---YDSISRA-----	282
Qy	167	PFVTVTNMGMNKIYYINGKLIDTIKVELTGTFNFKSTITFEINKIPDGLTITSDSDNIN	226
Db	283	YYCLVDTCANQM-----IEEYAGLNMDINLLVTFPFDAKSLPNQSSQLYDI-YIG	333
Qy	227	MWIRDYFIYAKELDGKD-----INTLFNSLOYTVVKDYWGNDIIRYN-KEYMYNIDY	278
Db	334	LWI-----DGLDEIFETREIVNSLFQYIGHKGQYKLUKITKGRNDLTENLIDEVASH	383
Qy	279	LNRYMYANSROI-----VFNTRRNNND-----FNEGYKILIKRIRGNTDNDFVRGGDI	326

Db 384 LNDLYHQEKKEISDVEDVIONKKETIIDIETVPFEEDLYSVISKLR----- 430  
QY 327 LYFDMTINNKAYNLFM-----KNETMYADNHSTEDIYAIGLREOTKDINDNI 373  
Db 431 VVVDLSLEPK---LFLQICCGAGIPQINKKRTDYVKH--MHNGYIIDDISQTVESLDYF 485  
QY 374 IFQIOPMNNYYIASQIFKSNFNGENI-----SGICSIGT 408  
Db 486 LAHLKNWNSYAYSML-TDDFSSINIHQINQLFKGDVSSGT 527

Search completed: November 7, 2002, 14:53:14  
Job time : 20 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 16:32:06 ; Search time 2721 Seconds  
(without alignments)  
4813.033 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPNIFSYNNSLKDI.....NYASLLESTHNGFVPVSE 450

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delcp 6.0 , Delcxt 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002\_111617\_20836/app\_query.fasta\_1.647  
-DB=GenEmbl -QFMT=fastp -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09910186@cgn\_1\_11616\_brunat\_04112002\_111617\_20836 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sv: \*  
13: gb\_un: \*  
14: gb\_yi: \*  
15: em\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_mu: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rnd: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2386	99.8	4479	7	CBCPHGCL X53751 Clostridium
2	2386	99.8	4592	7	CSTCITOX D90210 Bacterioph
3	2386	99.8	4712	1	CBPHNCIM X71126 C.botulinum
4	2386	99.8	9613	1	CBCITOX X62389 Botulinum b
5	2386	99.8	9689	1	CBP1CMTCL X66433 Clostridium
6	2386	99.8	11747	1	AB061780 X66433 Clostridium
7	2386	99.8	12297	7	CBCBONT X72793 Clostridium
8	1824	76.3	11631	1	AB037920 X8037920 Clostridi
9	1823	76.2	4049	1	CLOINCO D38442 Clostridium
10	882.5	36.9	11614	1	CLOBN D49440 Clostridium
11	882.5	36.9	11614	1	AB037166 X537166 Clostridium
12	867.5	36.3	3952	1	CBNTTD X54254 Clostridium
13	867.5	36.3	4087	7	S49407 X49407 type D neur
14	867.5	36.3	11584	7	AB012112 X49407 type D neur
15	610	25.5	3876	1	AF300465 X300465 Clostridi
16	610	25.5	3876	1	AF300466 X300466 Clostridi
17	604.5	25.3	3937	1	CBONTG X4162 C.botulinum
18	604	25.3	11170	1	CBY13630 X13630 Clostridium
19	602.5	25.2	3759	1	AB088207 X4162 C.botulinum
20	602.5	25.2	3838	1	AB039264 X4162 C.botulinum
21	602	25.2	3876	1	AF295926 X4162 C.botulinum
22	594.5	24.9	3759	1	AB082519 X4162 C.botulinum
23	594.5	24.9	4030	1	CBNTBE X26683 C.botulinum
24	594	24.8	3835	1	AB037704 X26683 C.botulinum
25	594	24.8	3835	1	AB037705 X26683 C.botulinum
26	594	24.8	3835	1	AB037706 X26683 C.botulinum
27	594	24.8	3835	1	AB037707 X26683 C.botulinum
28	594	24.8	3835	1	AB037708 X26683 C.botulinum
29	594	24.8	3835	1	AB037709 X26683 C.botulinum
30	594	24.8	3835	1	AB037710 X26683 C.botulinum
31	594	24.8	3835	1	AB037711 X26683 C.botulinum
32	594	24.8	3835	1	AB037712 X26683 C.botulinum
33	594	24.8	3835	1	AB037713 X26683 C.botulinum
34	594	24.8	3835	1	AB037714 X26683 C.botulinum
35	594	24.8	3949	1	CBNTBE X26683 C.botulinum
36	593	24.8	4051	1	CBONTB X26683 C.botulinum
37	586.5	24.5	3891	6	AR000031 X26683 C.botulinum
38	586.5	24.5	3891	6	AR169142 X26683 C.botulinum
39	586.5	24.5	3891	6	AR036248 X26683 C.botulinum
40	586.5	24.5	4292	1	CBOTAG X26683 C.botulinum
41	586.5	24.5	4835	1	CLONEUR X26683 C.botulinum
42	586.5	24.5	15039	1	AF461540 X26683 C.botulinum
43	583	24.4	4017	1	CBNEUTOXE X26683 C.botulinum
44	583	24.4	4017	6	AX088262 X26683 C.botulinum
45	581	24.3	4041	1	CLOBOTB X26683 C.botulinum

ALIGNMENTS

RESULT 1

CBCPHGC1  
LOCUS 4479 bp DNA linear PHG 12-SEP-1993  
DEFINITION Clostridium botulinum C phage gene for C1 neurotoxin.  
ACCESSION X53751  
VERSION 1  
KEYWORDS C1 neurotoxin; neurotoxin; toxin.  
SOURCE Clostridium botulinum C phage.  
ORGANISM Clostridium botulinum C phage  
Viruses.  
REFERENCE 1 (bases 1 to 4479)  
AUTHORS Popoff M.R.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1990) Popoff M.R., Institut Pasteur, Anaerobies,  
25 rue du Docteur Roux, 75724 Paris Cedex 15, France  
REMARK 2 (bases 1 to 4479)  
AUTHORS Hauser D., Eklund M.W., Kurazono H., Binz T., Niemann H.,  
Gill D.M., Boquet P., and Popoff M.R.  
TITLE Nucleotide sequence of Clostridium botulinum C1 neurotoxin  
MEDLINE Nucleic Acids Res. 18 (16), 4924 (1990)  
PUBMED 90370487  
REMARK 2204031  
AUTHORS revised by [3]  
REFERENCE 3 (bases 1 to 4479)  
AUTHORS Popoff M.R.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1990)  
FEATURES  
source  
1. 4479  
Location/Qualifiers  
/organism="Clostridium botulinum C phage"  
/strain="phage from 468C botulinum C. strain"  
/db\_xref="taxon:12448"  
198..203  
/note="pot. ribosome binding site"  
214..4089  
/note="C1 neurotoxin (AA 1-1291)"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAA37780.1"  
/db\_xref="GI:14906"  
/translation="MPITINNFSYSDVDRKNNILYDLTHLNLANEPEKAFRITGNIM  
VIPDRFSNPNLKNKPRVTSKSGYDNPYLSKSDKDPFLKEIKLFRKINSREI  
GEELYRLSTDPFPPGNNTPINTEFDFVDFNSVDVKTROCNWNNKTSINPSVITG  
PRENIDETSTFKLTNNFAOFGALSTISPREMLTYSNANDVGEGRFSKE  
KCMDDILLHELKHNHLYGIAIPNDQISSVTSNIFSOYNKLEAEIYAEQGP  
TIDLPSARKYFEKALDYRSIAKRLNITANPSFNKIYGEIKQKILKIRYV  
ESSGEVTVNRKVELYNELTQIEFENYAKIVNQNKIYLSNVYPTVNTAILDDNV  
YDIQNFNIPKSNLNVLFMGQNLNRPALRVNPNMLYLFKECHAKIDGRSLYNT  
LDCRELLVNTDLPFDIGDISDKTDIFLRKIDNEETEVYIPDNVSDQVILSKNTSE  
HGQDLVPSIDSSSEILPGENVQVYDNQNVYLSNYYLESQKLSNVDEFTFR  
STEEALDMSAKVTVYFPTLANKVAGQGLFLMWANDVDEFTNLRKTDKTSQ  
VSAITPYCPALNINSVVRGNTFEAFVGVITLLEAFPEFTIPALGAFVIYKVOE  
RNEILKIDNLEORIKRWKDSYEMWGMGLWSLRITOFNLSIQMDSLVNQAGAKA  
KIDLEBYKSGSDENKNSQVENLKSVDKISAMNNKIPRECSVYLFKNNLPK  
VIDELNEEDRTAKLINDSHIILVGEVDKLRKAVNNSPQNTIPFNLSYTNNSL  
LADIINEFNINDSKILSLQNRKNTLVDTSGYNAEYSEEGVQLAPIFPDKLGS  
GEDRKVTVINENISVNSYSEFISFWIRKNSVNLPCYTIIDSVKNSWSIGI  
ISNFTVFLKNESEQISNFSYDINNAPGYNKFFVTVTNNMGMKIYINGKLI  
TUKVELGAFSTKIPETINKIPOTGLIISDSNINMIRDEVIKAKELDGKIDIL  
FNSLYTVNVKDYDNGNDLRYNKEYMNYNIDLYNRYMANSRQIVFNTRRNNPNEGY  
KIIIRKNGNTNVRGGLIYFDMTINKNKYNLFKNNETMDADNHSIDETIAGLR  
EOTKDINPIYIQPMNTYIYASQIFKNSNGENISGICSTGYRFRUGGDWYRHN  
YLVPTVKQGNVASLESTSTHWGFVPVSE"  
BASE COUNT 1817 a 455 c 693 g 1514 t.  
ORIGIN

Alignment Scores:  
Pred. No.: 3,4e-152 Length: 4479  
Score: 2386.00 Matches: 449  
Percent Similarity: 100.00% Conservative: 0  
Bes% Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0

DB: 7 Gaps: 0  
US-09-910-186A-10 (1-450) x CBCPHGC1 (1-4479)  
QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21  
|||||  
Db 2740 ACAATACCCCTTTAAATATTTTTCATATACATAAATCTTTATTAAGAGATATAATTAAT 2799  
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
|||||  
Db 2800 GAATATTTCATATATTAATTAATGATTCAAATAATTTTGACCTACAAAACAGAAAAATACT 2859  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
|||||  
Db 2860 TTAGTGGATACATCAGGATATAATGCAGAACTGAGTGAAGAAGCGGATGTTACGCTTAAT 2919  
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
|||||  
Db 2920 CCAATATTTCATTTGACTTTAAATTTAGGTAGTTAGTTCAGGGAGGATAGGTAAGTTATA 2979  
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
|||||  
Db 2980 GTAACCCAGAAATGAAATATTGTATATAATTCATGATGAAAGTTTTCAGCATTAGTTT 3039  
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleLeuAspSerVal 121  
|||||  
Db 3040 TGGATTAGATAAATAAATGAGTAAATTTACCTGATATACATACTAATTAATGATAGTGT 3099  
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
|||||  
Db 3100 AAAATAACTCAGGTGGATAGTATAGTATTATTAGTAATTTTATTAGTATTACTTTAAAA 3159  
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161  
|||||  
Db 3160 CAAAATGAAGATAGTAGCAAAAGTATAAATTTTATGATATATCAATAATGATGCTCCT 3219  
QY 162 GlyTyrAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
|||||  
Db 3220 GGATACATAATATGCTTTTGTAACTCTTACTAACAATATGATGGGAAATATGAAGATT 3279  
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
|||||  
Db 3280 TATATAAATGSAATAATTAATAGATACTATAAAGTATAAAGTAACCTGGAATTAATTTT 3339  
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
|||||  
Db 3340 AGCAAACTATACATTTGAATAATAATAATTCAGATACCGGTTGATTACTTCAGAT 3399  
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
|||||  
Db 3400 TCTGATAACATCAATATGTTGGAAGAAGATTTTATATATTGCTAAGAATATAGATGGT 3459  
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261  
|||||  
Db 3460 AAAGATATTAAATATATTATTAAATAGCTTGCATATACATACTAATGTTGTAAGAATATTGG 3519  
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
|||||  
Db 3520 GCAAAATGATTTAAGATATAATAAAGAATATTATATGTTTAATATAGATTATTAAATAGA 3579  
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPhe 301  
|||||  
Db 3580 TATATGATGCGAAGCTCAGCAAAATTTGTTTATATACACGTAGAAATAATAATGACTTC 3639  
QY 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
|||||  
Db 3640 AATGAAGGATATAAATAATAAAGAAATCAGAGGAAATCAAGAGAAATGATGACTAGAGTA 3699  
QY 322 ArgGlyClyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
|||||  
Db 3700 CGAGGAGGAGATTTTATATTGATGCAATTAATAACAAGCATATAATTTGTTT 3759  
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361

```

Db 3760 ATGAAGATGAACATATGATCCAGATATCATCTAGTATGAGATATATATGCTATAGGT 3819
QY 362 LeuArgGluGlnThrLysAspLeuAsnAspLeuGlnLeuGlnProMetAsn 381
Db 3820 TTAAGAGACAACAAGATATAATGATATATATATATATATATATCAATACCAATGAAT 3879
QY 382 AsnThrTyrTyrAlaSerGlnLeuPheLysSerAsnGlyGluAsnLeuSer 401
Db 3880 AATCTATATATACCATCTCAATATATTAATCAATTTAATGGAGAAATATTTCT 3939
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421
Db 3940 GGAATATGTCATAGTACTTATGCTTTAGACTTGGAGTGATGATGATAGACAAAT 3999
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 4000 TATTTGGTCTACTGTGAAGCAAGAAATATGCTTCAATATTAGATCAACATCAACT 4059
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 4060 CATTGGGTTTGTACCTGTAAGTGAA 4086

RESULT 2
CSTCI TOX
LOCUS
DEFINITION Bacteriophage c-st (from C. botulinum) C1-tox gene for botulinum C1
neurotoxin.
ACCESSION D90210
VERSION D90210.1 GI:217780
KEYWORDS botulinum toxin; neurotoxin.
SOURCE Bacteriophage c-st (from C. botulinum type C-Stockholm) DNA.
ORGANISM
VIRUSES.
REFERENCE 1 (bases 1 to 4592)
AUTHORS Kimura,K., Fujii,N., Tsuzuki,K., Murakami,T., Indoh,T., Yokosawa,N.,
Yokosawa,N., Takashi,K., Suto,B. and Oguma,K.
TITLE The complete nucleotide sequence of the gene coding for botulinum
type C1 toxin in the C-ST phage genome
JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)
MEDLINE 9102498
REFERENCE 2 (bases 36 to 2143)
AUTHORS Kimura,K., Fujii,N., Tsuzuki,K., Murakami,T., Indoh,T., Yokosawa,N.
and Oguma,K.
TITLE Cloning of the structural gene for Clostridium botulinum type C1
toxin and whole nucleotide sequence of its light chain component
JOURNAL Appl. Environ. Microbiol. 57 (4), 1168-1172 (1991)
MEDLINE 91282468
COMMENT These data kindly submitted in computer readable form by: Kouichi
Kimura
Department of Microbiology
Sapporo Medical College
1 South, 17 West
Sapporo 060
Japan
E-mail: nfujii@niguts.nig.ac.jp&vax2.nlm.nih.gov
Phone: 011-611-2111 x2334
Fax: 011-612-5861

FEATURES
source
Location/Qualifiers
1..4592
/organism="Bacteriophage c-st"
/db_xref="taxon:12336"
255..260
/transl_table=1
293..298
/prot_id="BAA14235.1"
305..310
/db_xref="GI:217781"
321..4196
/translation="MPIINPNFNSDPVDNKNILYLDHNTLANEPEKAFRTGNLW
-note="SP sequence"
-note="ORF for C1 toxin"
/codon_start=1
/transl_table=1
/protein_id="BAA14235.1"
/db_xref="GI:217781"
/translation="MPIINPNFNSDPVDNKNILYLDHNTLANEPEKAFRTGNLW

```

BASE COUNT 1869 a 460 c 714 g 1549 t  
ORIGIN

# Alignment Scores:

3.48e-152 Length: 4592  
Pred. No.: 2386.00 Matches: 449  
Score: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 7 Gaps: 0

US-09-910-186a-10 (1-450) x CSTCI TOX (1-4592)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21  
Db 2847 ACATACCTTTAATATTTTTCATATCTATATAATCTTTTAAAGATATAATTAAT 2906  
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
Db 2907 GAATATTTCAATAATATTAATGATTCACAAATTTTTCAGCCTCAAAACAGAAAAATACT 2966  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2967 TTAGTGGATACATCAGGATATATGCGAAGTGAGTGAGAGGGGATGTCAGCTTAAT 3026  
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Db 3027 CCAATATTTCCATTGACTTTAAATAGGTAGTTCAGGGGAGGATAGAGTAAAGTTATA 3086  
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
Db 3087 GTAACCCAGAAATGAAATATTTGTATATAATCTATGTATGAAGTTTATGATAGTTT 3146  
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
Db 3147 TGGATTAGATAATAATAATGGGTAACTAATTTACCTGGATATATACTAATAATGATAGTCT 3206  
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
Db 3207 AAAAATAACTCAGGTGGAGTAGTAGTATATAGTAATTTTATAGTATTTTACTTTAAAA 3266  
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
Db 3267 CAAATGAAGATAGTAGACAAAAGTATAAATTTTAGTATGATATATCAATAATGCTCTCT 3326  
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
Db 3327 GGATACAATAAATGGTTTTTGTACTGTACTACAATATATGATGGAAATATGAGATT 3386  
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201

VIPREFSRNSPLNKPVRVTPSKSGYVDPNVLSTDSKDTFLKEIIKLFKRINSREI  
GEELIYRLSTDIPEPGNNTPINTDFDFDENSVDKTRGNNWWTGTSNPISVITIG  
PRENIIDPETSTFKLTNNTPFAQEGFALSTISISPREMLTYSNAINVLEGEFESKSE  
FCMDPILILHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLH  
TIDILPSARKYIEEKALDIYRSIAKRLNLTITANSPFNKYIGETKQKDLIRYFV  
ESSGEVTNRNKFVLYNELTQIFETFNKYAKYNQNRKYLSTNVPVPTANILDDNV  
YDQNGENIPKSNLNVLFMGQNLNRNPNALRVNPNMLYLFTECHKAIDRSLYKNT  
LDCRELVLKNDLPEIGDISDKVDIFLRKIDINEETEVYYPDNVSYDQVILSKNTSE  
HGOLDLYPSIDSESEILPGENOVYDNRTQNDYLSNYIYLSQKLSDNVEDFTFR  
SIEBALDMSAKVYTFPTLANKVNAGVGGGLFELMWANDVDEFTTNILRKDTLDKISD  
VSAIPIYGALNINSVVRGNFTEATVGTGVTILLFAFPEFTIPALGARVYIKVQE  
RNEIKITDNCLEQIRKWKDSYEMWTLRSITITQFNISQIMSDSLNQAQAKA  
KIDLEYKFGSGDKENIKSQVENLKSNDLVKISEAMNNKINFKIRECSVYLFKNMLPK  
VIDLENEFDRNTKALINLIDSHNIIIVGEVDKLKAKVNSFQNTIPFNITNNLSL  
LKDINEYFNNDNISKILSQNRKNTLVDTSQYNAEVSDEGDQVLPAPFFDFKLSG  
GDKRGVIVTONENIVNVSYESISFWIRINKWNSNLPYITIDSVKNSWSGICI  
ISNELYFTLKQNEDESEINSEYDISNAPGYNKWFVTVTNMNMKMTIYINGKIL  
TIKVELGIGNFSKTIIFENKIDPTGLITSDSDNINMWIRDEYFAKEIDGDKJNL  
FNSLOYTNVVKDYNGNDLRNKNYDIYLNRIYANRSQIVFVTRRNNDFNEG  
KIIIRKIRGNTDTRVRGGDILYFDFTINNKAYNLFMKNMETMYADNHSTEDIYALGR  
BOTKDINDNIIFOIQPMNTIYASQIFKSNFNGENISGICSGTGYFRFLGSDWYRHN  
YLVPTVKOGNVASLLESTSTHWGFPVPE"

Db	3387	TATATAAATGGAAATTAATAGATACTATAAAAGTTAAAGAACTAACTGAATAATTTT	3444
QY	202	SerLysThrIleThrPheGluIleAsnLysIleProaspThrGlyLeuIleThrSerAsp	221
Db	3447	AGCAAACTATAACATTGAATAATAAATCCAGATACCGGTTTGATTACTTCAGAT	3506
QY	222	SerAspAsnIleAsnMetTrpIleArgaspPheTyrIlePheAlaLysGluLeuAspGly	241
Db	3507	TCGTGATAACATCAATATGTTGGATAAGAGATTTTATATATTGCTGAAGAAATTAGATGGT	3566
QY	242	LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp	261
Db	3567	AAAGATATTAAATATATTATTATACCTTGCATATATACTATGTGTGAAGAATATTGG	3626
QY	262	GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValasnlleAspTyrLeuAsnArg	281
Db	3627	GGAAATCATTTAAGATATAATAAAGAATATTATATGTTTAATATAGATATTATTTAAATAGA	3686
QY	282	TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAspPhe	301
Db	3687	TATATGTTATGCGAACTCAGACAAATGTGTTTTAATACACGTAGAATAATAATGACTTC	3746
QY	302	AsnGluGlyTyrIlyIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal	321
Db	3747	AATGAAGGATATAAAATTAATAATAAAGAATCAGAGAAATACAAATGATACTAGAGTA	3806
QY	322	ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe	341
Db	3807	CGAAGGAGGAGATATTTTATTTTCATATGACAAATAATAACAAGCATATAATTTGTTT	3866
QY	342	MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly	361
Db	3867	ATGAAGAATGAAATCATGTATGCAGATAATCATAGTACTGAAGATATATGCTATAGGT	3926
QY	362	LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn	381
Db	3927	TTAAGAGACAAACAAGGATATAATGATAATATATATTTCATAATACAAACCAATGAAT	3986
QY	382	AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer	401
Db	3987	AATACATTATATTACGCATCTCAATATTTAATCAATTTTAAATGAGAAATATTTCT	4046
QY	402	GlyLecSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn	421
Db	4047	GGAAATATGTTCAATAGTACTTATCGTTTACGTTGAGGTGATTTGGTATAGACAAAT	4106
QY	422	TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr	441
Db	4107	TATTTGGTGCCCTACTGTGAAGCAGAGAAATTATGCTTCATTATTAGAAATCAATCACT	4166
QY	442	HisTrpGlyPheValProValSerGlu	450
Db	4167	CATTGGGGTTTGTACCTGTAAGTGAA	4193
RESULT 3			
CBPHNCNIM			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			

TITLE	Comparative analysis of C3 and botulinum neurotoxin genes and their environment in Clostridium botulinum types C and D
JOURNAL	J. Bacteriol. 175 (22), 7260-7268 (1993)
MEDLINE	94042900
PUBMED	8226673
FEATURES	Location/Qualifiers
source	1..4712
	/organism="Clostridium botulinum"
	/strain="type C (non toxic)"
	/sub_strain="A028-CN phage"
	/db_xref="taxon:1491"
RBS	791..796
gene	807..4682
	/gene="neurotoxin mutated gene"
CDS	807..4682
	/gene="neurotoxin mutated gene"
	/pseudo
	/codon_start=1
	/transl_table=11
	/db_xref="PID:e9305"
misc_feature	2463..2465
	/gene="neurotoxin mutated gene"
	/note="intragenic stop-codon"
BASE COUNT	1907 a 482 c 721 g 1602 t
ORIGIN	
Alignment Scores:	
Pred. No.:	3.57e-152 Length: 4712
Score:	2386.00 Matches: 449
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	99.79% Indels: 0
DB:	1 Gaps: 0
US-09-910-186a-10 (1-450) x CBPHCNCLM (1-4712)	
Qy 2	ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21
Db 3333	ACAATACCCTTTAAATATTTTTCATATACTAATAATTCCTTTATTAAGATATAATTAAT 3392
Qy 22	GlutYrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 3393	GAATATTTCAATAATTAATGATTCAAAATTTTGAGCTACAAACAGAAAATACT 3452
Qy 42	LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 3453	TTAGTGGATACATCAGGATATAATGCAGACAGTGAAGAACGCGATGTTCAGCTTAAT 3512
Qy 62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 3513	CCAATATTTCCATTTGACCTTTAAATTAGGTAGTTAGTGGGGAGGATAGAGTAAAGTTATA 3572
Qy 82	ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 3573	GTAAACCCAGAAATCAAAATATGTATATTAATTCATGATCAAGAGTTTACGATTAGTTT 3632
Qy 102	TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleLeuAspSerVal 121
Db 3633	TGGATTAGAATAAATAAATGGTAGTAATTTTACCTGGATATCTATAATTGATAGTGT 3692
Qy 122	LysAsnAspSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 3693	AAAAATACCTCAGGTGGAGTATAGGTATATTAGTAAATTTTATGATTACTTTTAAAA 3752
Qy 142	GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161
Db 3753	CAAAATGAAGATAGTGAACAAAGATATAAATTTTACTTATGATATATCAATAATGCTCT 3812
Qy 162	GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3813	GGATACAAATAATGGTTTTTGTACTGTACTAACAAATATGATGGGAATATGAAGATT 3872
Qy 182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuLeuThrGlyIleAsnPhe 201

Db 3873 TATATAAATGGAATTAATGATCTATAAAGTAAAGAACTAACTGGAATTAATTT 3932  
 QY 202 SerLysThrIleThrPheClnrleasnLysIleProaspThrGlyLeuIleThrSerAsp 221  
 Db 3933 ACRAAACTATPACATTTGAATAATAATAATTCAGATACCGTTTGATCTACAT 3992  
 QY 222 SerAspAsnIleAsnMetTrpIleArgaspPheTrpIlePheAlaLysGluLeuAspGly 241  
 Db 3993 TCTGATAACATCAATATGCGATAGAGATTTTATATATTGCTAAAGAAATAGATGGT 4052  
 QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnrThrAsnValValLysAspTrpTip 261  
 Db 4053 AAGATATTAATATATTTAATAAGCTGCAATATCTAATATGTTGTAAGAGATATTGG 4112  
 QY 262 GlyAsnAspLeuArgTrpAsnLysGlnrTrpMetValAsnIleAspTrpLeuAsnArg 281  
 Db 4113 GGAATGATTTAAGATATAATAAAGAAATATTATATGTTAATATAGATTTTAAATAGA 4172  
 QY 282 TyrMetTrpAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAspPhe 301  
 Db 4173 TATATGATGCGAATCGACAAATGTTTTTAAATACACGTAGAAATAATAATGACTTC 4232  
 QY 302 AsnGluGlyTrpLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 Db 4233 AATGAAGATATAAATTAATAAAGAAATCAGAGAAATACAAATGATCTAGAGTA 4292  
 QY 322 ArgGlyGlyAspIleLeuTrpPheaspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
 Db 4293 CGAGGAGGAGATTTTATTTGATGATGCAATTAATAACAAGATATAATTTGTTT 4352  
 QY 342 MetLysAsnGlnrThrMetTrpAlaAspAsnHisSerThrGluAspIleTrpAlaIleGly 361  
 Db 4353 ATGAAGAAATGAACATGATGACGATATCATCTACTGAGATATATATGCTATAGGT 4412  
 QY 362 LeuArgGlnrThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 Db 4413 TTAAGAGAAACAACAAGATATAAATGATATATATATTTCAATACACCAATGAAT 4472  
 QY 382 AsnThrTrpTrpTrpAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 Db 4473 AATCTATATATAGCATCTCAATATTTAATCAATATTTATGAGAAATATTTCT 4532  
 QY 402 GlyIleCysSerIleGlyThrTrpArgPheArgLeuGlyGlyAspTrpTrpArgHisAsn 421  
 Db 4533 GGAATATGTTCAATAGGATCTATCTGTTTATGACTTGGAGGTGATGTTGATAGACAAAT 4592  
 QY 422 TyrLeuValProThrValLysGlnrGlyAsnTrpAlaSerLeuLeuGluSerThrSerThr 441  
 Db 4593 TATTGGTCCCTACGTGAGCAAGAAATTAATGCTTCTATTAGATCAACATCAACT 4652  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 Db 4653 CATTTGGGTTTGTACCTGTAAGTGA 4679  
 RESULT 4  
 LOCUS CBCTOX 9613 bp DNA linear BCT 29-JUN-1999  
 DEFINITION Botulinum bacteriophage genes for HA-17, HA-33, nontoxic components and C1 toxin.  
 ACCESSION X62389  
 VERSION X62389.1 GI:558175  
 KEYWORDS botulinum toxin; C1 toxin; HA-17 gene; HA-33 gene; Hemagglutinin; neurotoxin; nontoxic; toxin.  
 SOURCE Clostridium botulinum.  
 ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
 REFERENCE 1 (bases 1 to 9613)  
 AUTHORS Kimura, K., Fujii, N., Tsuzuki, K., Murakami, T., Indoh, T., Yokosawa, N., Takeshi, K., Suto, B. and Oguma, K.  
 TITLE The complete nucleotide sequence of the gene coding for botulinum type C1 toxin in the C-ST phage genome

JOURNAL MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source  
 gene  
 CDS  
 RBS  
 gene  
 CDS  
 RBS  
 gene  
 CDS

Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)  
 91024998  
 2222445  
 2 (bases 1 to 9613)  
 Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N. and Oguma, K.  
 Nucleotide sequence of the gene for one of the components of hemagglutinin produced by Clostridium botulinum type C unpublished  
 3 (bases 1 to 9613)  
 Tsuzuki, K.  
 Direct Submission  
 Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept of Microbiology, South 1 West 17, Sapporo 060, JAPAN  
 On Oct 15, 1994 this sequence version replaced gi:40386.  
 See also X53041 & D90210.  
 Location/Qualifiers  
 1. 9613  
 /organism="Clostridium botulinum"  
 /strain="type C Stockholm, C-st"  
 /db\_xref="taxon:1491"  
 /complement(111..551)  
 /gene="HA-17"  
 /complement(111..551)  
 /gene="HA-17"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44260.1"  
 /db\_xref="GI:40387"  
 /translation="MSERTFLPKNYKSLFSDSLYITYSGALSFNTSLNOK WKLEYISNGFRFSNVAEPKYLAYNDYGFYLSSSNNLWNPKIAINSYIICL SIVNVDYAWTIYDNNNNITDQPIILNPLNPDINNNOILKLEK"  
 complement(560..565)  
 complement(614..1474)  
 /gene="HA-33"  
 complement(614..1474)  
 /gene="HA-33"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44261.1"  
 /db\_xref="GI:40388"  
 /translation="MSQTANDLRNNEVFIFSPSNNTNKVLDKISOSEVKLNKLSGA NKWRLIYDNKQAYKIKVMDNTSLITWNPASSVSKDITNGDQNTWLLQNIYSR NVIRYNTNPNVLQYNIIDTLMVSTQSSNQFFKFSNCIYEALNNRCKLQTLNS DRFSKLNQIYVLWQWPSRQKWIIEYETKSAITKQENRYLWQNNYIV ETYQSDSLIQYWNINLDNASKYILYLNQDTRVLDVNSQIANGTHVIVDSYHGN TNOQWITNL"  
 complement(1482..1486)  
 1726..1731  
 1737..5327  
 1737..5327  
 /gene="nontoxic"  
 /gene="nontoxic"  
 /codon\_start=1  
 /transl\_table=11  
 /product="non-toxic components"  
 /protein\_id="CAA44262.1"  
 /db\_xref="GI:40389"  
 /db\_xref="SWISS-PROT:P46081"  
 /translation="MDINDLNTNSPYDNKNVYIVRARKNTFFKAFKVPNIWAPE RVGPELDIAEYKLDGGIYDSNPLSODSERENFLOAIILLKRINNTISGKOLLSLI STATPPFVGITGGYSSPNIPTFGKTPKSNKLSLVSTIPFPFGYRETNYESQN NKFYASNIIFGPGSNIVNNVYKNDANGMTMAEIPQPLLYKYNKYIDP AWELTKLISLFIIGIKESDNLVWPYRLRTELDKQFSQNLIDLLISGGVDLEI NTNPWTNSYFPNSIKMFYKNIYKTEIEGNAIGNDKLKRQFQINQVODIWNL NLNTFOSFNSIIPDRSNALKHYRKYQYTDYTDNYINGFVNGQINTKPLSNKN TNIISKPKVNLVNNENIILMKSNIYDGKGTEDFYSTYKIPYNEEYEFNDS TNNENFISDPSIPEIIDIINPKNSDNLVETQITSMTEETITALSINYLQAI NFPLNFISSDPSKSVSSKSLVSPFLDNLMSYLETKNDGPIDTDKYLKKEVF KNSYFDNLQELDSMGINVWLPKALNLTNSFVEEYODSALSISKKNL REPNIIDDSIDSLGSLGNKNIYKALNLTNSFVEEYODSALSISKKNL LCMWAKOSILAQESLVKQIVONKFTDLKSLIPDITLKLIRETEKTFIDLSNESQIS

MSRVDNLFNKAISCVFEDYIPKFIYSIMEKYINNNIKTRFIOQCTNINLNEKSILL  
NSVFTKIDFKFDIOSIKNFNSQVEQWKEILSPYQLLLFASKGPNNSIIEIDSK  
NTLIQVTESELVYGVNGSLIKSPNEIIKFSNKFNFINGLNFNGFIFWRFKQND  
KDFLGNKVNNGCWGIIYEDNGLVEFIIDISNGNESVYLSNINNNWYIISISVDR  
KDQLLFINDKNVANSIQIUISYIETNINSLVKNNSIYVEELSVLDPNPISEVR  
NYFSYLDNSYRSDSSLLYNNKLYNYVPEPTESLYKNDNNKYSLSKNTQDINI  
SVYFKLINIDESKTVQWDECIICVLDGTEKYLIDISPNRNRIOLVSKNNAKITV  
NTDLFRPDCITFSYNDKYFSLSLRDGDYNNMICNDNNKVPKGAHLWLES"  
5330. 5335  
RBS  
5342. 9217  
gene  
/gene="Cl toxin"  
5342. 9217  
CDS  
/gene="Cl toxin"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA44263.1"  
/db\_xref="GI:40390"  
/db\_xref="SWISS-PROT:PI8640"  
/translation="MTITNNFSDVFNKMLIKLDHLNLANEPKAFRTGNW  
VIPDRSRNPNLKNRPRTSPKSGYDNPYLSLSDSKDTFLKEIKLFRINSREI  
GELIYRLSDIPFGNNPTINFDYDFNSVDVKTROGNNWKTGINSPIVITG  
PRENIIDPSTFKLNNTPFAQEGFALSIISSIPNMLTYSNATNDVGRFSKE  
FCMDPILMLHMLNHAHNLGYTAIPNDQTISSVTSNIFYSOYNKLEYAEIYAFGP  
TIDLIPKSARKYFEERLADYYSIAKELANSITANPSSFNKYIGYKOKLIRKPEV  
ESGEVTVNRKKEVELNLTQIFTEFNAYKIYNQNRKILSNWITPYANILDDNV  
YIQONGNPIKPSNLVFWQNLNRPNALRVNPNMULYLFKFKHKAIDGRSLYNT  
LDCRELKNTDIPFGIDSDVKTDFLKRKINEETEVYIPDNYSVDQVILSKNTE  
HGOLDLPISIDSEILPGENQVFDNRTQNVYINSYVLESOKLSNVEDFTFR  
SIEPALDNSKAVVYPTPLANKVNGVGGFLMWANDVEDFTNLRKOTLDKISD  
VSAIPIYGALNINSVRNGNEFAVGTVITLEAPPEPTIPALGAFVYKVOE  
RNEILWIDNCLPORKRWKDSYEWMMGTWLSRIITQENNISYOMVDSLNOAGAKA  
KIDLEKYSKSGSKENIKSQVENKSLDSEAMNNKINRECSVYTFYFNKMLP  
VIDELNFRDNTAKLINIDSHNIIIVGEVDKLRKAKVNNNSPONTIPNIFSNLSP  
LKQITNEFVNINDSKLSONKNTLVDTGYNAEVSEGVQNLNIPFPDFKLGG  
GEGRKVIYTONENIVNYSFISFVIRINKVSNLPVTIIDSVKNSGWSIGI  
ISNPLVLTGNEKSDQSINFSYDINSNAPGVNKFVTVTNMKNKIIYINGKID  
TIKVELTGNFESKTIIFELNIPDGLITSDSNINMWIRDFYIFAKELGKIDNII  
FNSLOLVNVDKWDGNDLRNKEYVWYNDLNRKYANSRQIVENTRNNDNEG  
KLIKIRGNTNTRVGGDILFDPMNNKAYINLFMKNETIADNHSDEIYATGLR  
EQNDINDNIIIFIQPMNNITYIASQIFKSNFNGENISGICSIGTYRFRFGDWRHN  
YLVPYKQGNYSALLESTTHWGFVPSV"

BASE COUNT 3846 a 1003 c 1348 g 3416 t  
ORIGIN

Alignment Scores:

Pred. No.: 7 23e-152 Length: 9613  
Score: 2386.00 Matches: 449  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 1 Gaps: 0

US-09-910-186a-10 (1-450) x CBCTOX (1-9613)

Qy	2	ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleLeuAsn	21
Db	7868	ACAATACCTTTAATATTTTTCATATACTAATAATCTTTTATAAAGATATAATTAAT	7927
Qy	22	GlutTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
Db	7928	GAATATTTCATATAATTAATGATTCAAAATTTTGGCCCTACAAACAGAAAAATACT	7987
Qy	42	LeuValAspThrSerGlyTyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn	61
Db	7988	TTAGTGGATACATCAGGATATAATGCAGAGTGTGAGTGAAGAGCGGATGTTACGCTTAAT	8047
Qy	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluaspArgGlyLysValIle	81
Db	8048	CCAAATTTCATTTGACTTTAAATAGGTAGTTTCAGGGGAGGATAGAGTAAAGTTATA	8107
Qy	82	ValThrGlnAsnGluAsnIleValTyTyrAsnSerMetTyTyrGluSerPheSerIleSerPhe	101
Db	8108	GTAAACCCAGAAATGAAATATGTTATATAATCTATGATGCAAGAGTTTACCATAGTTT	8167

Qy	102	TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleaspSerVal	121
Db	8168	TGGATTAGAATAAAATAATGGTAAGTAATTTACCTGGATATACATAATGATAGTGT	8227
Qy	122	LysAsnAsnSerGlyTyTyrPsrIleGlyIleIleSerAsnPheLeuValPheThrLeuLys	141
Db	8228	AAAAATAACTCAGGTGCGAGTATAGGTATTATTAGTAATTTTTTAGTATTACTTTAAAA	8287
Qy	142	GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyTyrAspIleSerAsnAsnAlaPro	161
Db	8288	CAAAATGAAGATAGTAGCAAAAGATATAAAATTTTAGTATGATATATCAAAATATGCTCT	8347
Qy	162	GlyTyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
Db	8348	GGATACAAATAAATGGTTTTTGTAACTGTTTACTAACAAATATGATGGAAATATGAAGATT	8407
Qy	182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe	201
Db	8408	TATATAATGGAATAATTAATAGATCTATATAAAGTTAAAGAACTAACTGGAATTAATTTT	8467
Qy	202	SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp	221
Db	8468	ASCAAACTATACATTTGAAATAAATAAATTTCCAGATACCGGTTTGATTACTTCAGAT	8527
Qy	222	SerAspAsnIleAsnMetTrpIleArgAspPheTyTyrIlePheAlaLysGluLeuAspGly	241
Db	8528	TCTGATAACATCAATATGCGATAAAGATTTTTTATATATTTCTCTAAAGAATTAGATGT	8587
Qy	242	LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyTyrTyr	261
Db	8588	AAAGATATATATATATATTTTATAGCTTCCAATATACATACTTAATGTTCTAAAGATATTGG	8647
Qy	262	GlyAsnAspLeuArgTyTyrAsnLysGluTyTyrMetValAsnIleAspTyTyrLeuAsnArg	281
Db	8648	GGAAATGATTTAAGATATAATAAAGAATAATATATGTTAAATATAGATATTATTAATAAGA	8707
Qy	282	TyrMetTyTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe	301
Db	8708	TATATGATGCGAACTCACGACAAATTTGTTTTTAACACGTAGAAATATATATGACTTC	8767
Qy	302	AsnGluGlyTyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal	321
Db	8768	AATGAAGATATAAAATTTATAATAAAGAAATCAGAGGAAATACAAATGATACATAGATA	8827
Qy	322	ArgGlyGlyAspIleLeuTyTyrPheAspMetThrIleAsnAsnLysAlaTyTyrAsnLeuPhe	341
Db	8828	CGAGGAGGAGATATTTTATATTTTGTATGACCAATTAATAACAAAGCATATAATTTGTTT	8887
Qy	342	MetLysAsnGluThrMetTyTyrAlaAspAsnHisSerThrGluAspIleTyTyrAlaIleGly	361
Db	8888	ATGAAGATGAACATATGATGCAGATAATCATAGTACTGAGATATATATGCTATAGCT	8947
Qy	362	LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn	381
Db	8948	TTAAGAGAACAACAAGAGATATAAATGATAATATATATATATTTCAAAATACAAACAAATGAAT	9007
Qy	382	AsnThrTyTyrTyTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer	401
Db	9008	AACTATTATTTACGCATCTCAAAATTTTAAATCAAAATTTTAAATGGAGAAATATTTCT	9067
Qy	402	GlyIleCysSerIleGlyTyTyrArgPheArgLeuGlyAspThrTyTyrArgHisAsn	421
Db	9068	GGAATATGTTCAATAGGTACTTATCGTTTGTAGACTTTGGAGGTGATTGGTATAGACACAAT	9127
Qy	422	TyrLeuValProThrValLysGlnGlyAsnTyTyrAlaSerLeuLeuGluSerThrSerThr	441
Db	9128	TATTGGTCCCTACTGTGAAGACAGGAAATATGCTTCATTATTAGAAATCAACATCACT	9187
Qy	442	HistPheGlyPheValProValSerGlu	450
Db	9188	CATTGGGGTTTTGTACCTGTAAGTGAA	9214

```

RESULT 5
CBP1CMT1  CBF1CMT1  Clostridium botulinum phage 1C, CHn-14, CHn-33, CHn-138 and Bont/C1
DEFINITION 9689 bp DNA linear BCT 15-FEB-1999
LOCUS
X66433.1 GI:509271
ACCESSION X66433.1
VERSION haemagglutinin; neurotoxin.
KEYWORDS Clostridium botulinum.
SOURCE Clostridium botulinum.
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
          Clostridium.
REFERENCE 1 (bases 1 to 9689)
AUTHORS Hauser,D.F., Eklund,M.W. and Popoff,M.R.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 9689)
TITLE Direct Submission
JOURNAL Submitted (23-MAY-1992) D.F. Hauser, Institut Pasteur, Unite des
          Toxines Microbiennes, Institut Pasteur, 28 rue duDocteur-Roux,
          F-75724 Paris Cedex, FRANCE
FEATURES
source 1..9689
         /organism="Clostridium botulinum"
         /db_xref="taxon:1491"
         complement(253..647)
         /gene="Chn-14"
         complement(253..633)
         /gene="Chn-14"
         /codon_start=1
         /transl_table=11
         /product="14kDa protein associated with Bont
         /C1-haemagglutinin complex"
         /protein_id="CAA47057.1"
         /db_xref="GI:509272"
         /db_xref="SPRMBL:045915"
         /translation="MSSERTFLPNKYKKSLEFSDSLYTYSSGALSPFNTSSLDNQK
         WKLYISISSGFRFSNVAEPKNKYLAINDYGFYLSSSNSLNWPKIAINSYIICTL
         SIVNTDYATVIYDNNNNIIDHLF"
         complement(642..647)
         /gene="Chn-14"
         complement(696..1556)
         /gene="Chn-33"
         complement(696..1556)
         /gene="Chn-33"
         /codon_start=1
         /transl_table=11
         /product="33kDa protein associated with Bont
         /C1-haemagglutinin complex"
         /protein_id="CAA47058.1"
         /db_xref="GI:509273"
         /db_xref="SPRMBL:045916"
         /translation="MSQTNANDLRNNEVFFISPSNNTNKLVDKISQSEVKLNKLSGA
         NQWRLLYDNTKQAYKLYMDNTSLITWNAPLSSVSKTDTNGDQYWLQNYISR
         NVIIATNYPNLYQYIDIDTLAVSTQTSNNOFFKFSNCIYEALNRRNKLOTLNS
         DRFLSKNLGSIYLVQWOFDSRQKWIIEYNETKSATLKCOENRRLTWIQNSNYY
         ETVQSTSLQYWNINYLNDASKYILYNQDTRNRLVDVNSQIANGTHVIVDSYRGN
         TNQQWIIINLI"
         complement(1564..1568)
         /gene="Chn-33"
         complement(1564..1568)
         /gene="Chn-33"
         join(1803..1808,1819..5409)
         /gene="Chn-138"
         /gene="Chn-138"
         /gene="Chn-138"
         /codon_start=1
         /transl_table=11
         /product="138kDa protein associated with Bont
         /C1-haemagglutinin complex"
         /protein_id="CAA47059.1"
         /db_xref="GI:4376225"

/translation="MDINDNLINSVDNKNVIVIVARKTIEFKAKVAPNIVAPE
RYTEPDLDAEEYKLDGGIYDSNLSQSDSERENFLOAIILLLKRNINISGKQLLSLI
STAIPFYGGYISPNITFTGKTPKSNKNLNSLVSTIIPFPGGGRETNYSIESON
NKFYASNVLIIFGSGNIVENNVIYKNDANGMGTMATVIFPQGLLYKYNKFYIDP
AMELTCLIKSLYLYGIRKPSNDLVYPRTELNDKQFQSNLIIDLLISGVDLEFI
NTNPFNTSYSPFNKPEKYNKYKTEIEGNAIGNDIKRLKQKQFQINQVQDNKNL
NLAVFCQSPNSIIPDRFSNALHFYKQYTYDNTDYNINGFVNGQINIKLPUNKN
TNIIISPEKVVNLVNNNISLMSKNLYGDLGKTGTFDYSTVKIPNEEVEYRNDSD
NFPLNISIEEVSIPEDIDINPYKONSNLVFTQITSTWETVTHALISNLAQI
TNNEFTLSDFSQVSSKDKSLVYFLDNLSYLETIRKNDGPIDHDIYILWLKVEF
TNFNIDTQIEDSMCGNEVYVWFGKALINLNTSNSEVEYDSGSA.SLSLKDNL
REFNIEDIDISLLGSLFNDLNNKLYISKNIVFYFKYFSDQWMTYYSOYFE
LIMAKOSILAQESLVKQIVQNKFTLSKASIPPDPTKLIRETEKTEFLDNESQIS
MNRVDNFKLASICVFEVDIYKFIYSMEKYNINNIKRETIQCTINIDNEKISLI
NSYFTKIDPFKLDIOSIKNFNSQVQMKRILSPYQLLPASGPNNSNIIDTSK
NTLIQTESIELYGVNGESLILKSPNEIKFSNKEFTNGLTNCTICFWLRETKGND
DKTRLGKNNVNCWEIIFEDNGLVFEIISNGNSESYSLSNIDNWWYISISVDRL
KQDILLINDKNVANYSIQISYTNISLVNKNNSIYVEELSVLNDNPTSEVIR
NYFSYLDNSYIRDSKSLLEYKNKYLYNVPFETSLYEVDNNSKYSLSLNTGINI
SSYFKLINIDESKVVYQKWDECIICVLDGTEKYLIDISPNNRILQSVSKNAKKITY
NTDLFEDPCITFSYNDKFKYLSLSDGDYNNWMCNDNNKVPKGAHLWILES"
5408..5413
/gene="Bont/C1"
5408..5413
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
/codon_start=1
/transl_table=11
/product="Botulinum C1 neurotoxin"
/protein_id="CAA47060.1"
/db_xref="GI:509275"
/db_xref="SWISS-Prot:P18640"
/translation="MPTTNNNSYSDVDNKNILYDLTHLNTLANEPEKAFRITGNIV
VIDRPSRNSPNLNPAPRVTSKSPSYDPNVLSDSDKDTKLKILKFKINSREI
GEELIYLSIDIPFGNNNPINIFDQVDFNSVDKTRQGNWVKTSINFSVLTIG
PRENIDPSTFKLNTNTPAAGEGFALSIIISPRMLTYSNATNDVWGGRFSKSE
FCMDPILILMHELHAMNLYGIAIPNDQTIISVTSNFIYSQYNKLEYAEIYAFGGP
TIDLPKARKYFEKALDYRSIAKLSITANPSFNKYGKQLIKRYFRVY
ESSGEVTNKNFVELYNELTQITFTFNKYNQNRKILYLSNVTPTANILDDNV
YDQNGNIPKSNLNVFMGONLSRNPALKRVNPNMLYFTKCHKAIDGRSLNKT
LDGRELLYKNTDLPFTGDISDKVDIPLSKDNEETEVIIYPDNYSVQVILSKNTE
HGQDLIYPSIDSESLPGENOVYDNRQNDVILNSYXLESQKLSNDVEDFTFR
SIEALDNSAKVITYPTLANKNAGVQGLFLMWANDVVEFTNLKDKDLKISD
VSAIIPYPALNISNSVRGNFTEAPVGTVILLEAPEPTIFALGAFVIYSKVE
RNEIITIDNCLQRIKRWKDSYEMWGTLSRIITQFNISIQYDLSNYQGAIGA
KIDLEYKYSQSKENIKSQVENKSLDVKIASEMNNINKFIRECSYTLFKNMLPK
VIDELNEFDRTAKLINLQNRKNTLVDTSGVNAVSEBQVQLNPFPFDFKLGSS
LKDIINEFVNINDSKILSNKNTLVDTSGVNAVSEBQVQLNPFPFDFKLGSS
GEGRGKVIYQNNENIYNSMESFSISFWIRINKVSNLPGVTIIDSVKNSGWSIGI
ISNPLVFTLQNSDESQSINFSDISNNAPGNKFFVIVNMMGNMKIYINGKILID
TIKVKELTGINFSTKITEINKIPDTGLTSDSDNINMWIRDFYIFAKELQKDNIL
FNSLQYTNVYVYWGNDLRYKKEYVNDIYNRYMYANSROIIVNTRNNDFNEGY
KIIIRIRGNTDTRVGGDILYFDMTINNKAYNLFPMKNETYADNHSIEDIATGLR
EQDKINDNTIIFOQPMNNTYVYASQIFKSNFNGENISGICSTGTFRFLGGDWVRHN
YLVPTVKQGNVASLLESTSTHWGFVPVSE"
9385..9424
/gene="Bont/C1"
9385..9424
/gene="Bont/C1"
/note="putative"

BASE COUNT 3865 a 1022 c 1359 g 3443 t
ORIGIN

Alignment Scores:
Pred. No.: 7,29e-152 Length: 9689
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
Db: 1

```



D	b	8970	ATGAGAAATGAACCATATGTATGCAGATTAATCATATAGTACTGCAAGATATATATCGTTATAGT	9029
Q	y	362	LeuArgGluGlnThrIleAspIleAsnAspAsnIleIlePheGlnIleGlnPrometAsn	381
D	b	9030	TTAAGAGAACAACAAGAGATATAAATGATTAATATTATATTCAAAATACAACCAATGAAT	9089
Q	y	382	AsnThrTyrtYrtyrAlaserGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer	401
D	b	9090	AATACCTATATATAGCCATCTCAATATTAAATCAAAATTTTATATGGAGAAATATTTCT	9149
Q	y	402	GlyIleCysSerIleGlyThrTyrtArgPheArgLeuGlyGlyAspTrpTyrtArgHisAsn	421
D	b	9150	GGAATATGTTCAATAGTACTATTCGTTTAGACTTGAGAGTGGATGGTATAGACACAT	9209
Q	y	422	TyrtLeuValProthrVallyslGlnGlyAsnTrpAlaserLeuLeuGluSerThrserThr	441
D	b	9210	TATTTGGTGCCTACTGTGACGACGAGGAATATATCTTCATATTAGAACATCAATCACT	9269
Q	y	442	HistrrGlyPheValprovalserGlu	450
D	b	9270	CATGGGGTTTTGTACCTGTAGTGAA	9296
R	E	S	RESULT 6	
L	O	C	AB061780	
D	E	F	LOCUS	
D	E	F	DEFINITION	AB061780 11747 bp DNA linear BCT 06-NOV-2001
A	C	C	ACCSSION	Clostridium botulinum orf-22, ha-70, ha-17, ha-33, ntnha, nt genes
V	E	R	VERSION	AB061780
K	E	Y	KEYWORDS	AB061780.1 GI:16580753
S	O	U	SOURCE	Clostridium botulinum (strain:C-Yoichi) DNA.
S	O	U	ORGANISM	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
R	E	F	REFERENCE	Clostridium.
A	U	T	AUTHORS	1
T	I	T	TITLE	Sagane,Y., Kouguchi,H., Watanabe,T., Sunsgawa,H., Inoue,K., Fujinaga,Y., Oguma,K. and Ohiyama,T. Role of C-terminal region of HA-33 component of botulinum toxin in hemagglutination Biochem. Biophys. Res. Commun. 288 (3), 650-657 (2001)
J	O	U	JOURNAL	21534265
M	E	D	MEDLINE	2 (bases 1 to 11747)
R	E	F	REFERENCE	Sagane,Y., Watanabe,T., Kouguchi,H., Morita,M., Miyamoto,T., Yamamoto,T., Muroga,A. and Ohiyama,T.
A	U	T	AUTHORS	Direct Submission
T	I	T	TITLE	Submitted (16-MAY-2001) Toshihiro Watanabe, Tokyo University of Agriculture, Faculty of Bioindustry; Yasaka 196, Abashiri, Hokkaido 099-2493, Japan (E-mail:t-watana@bioindustry.nodai.ac.jp, Tel.81-152-48-3843, Fax:81-152-48-3843)
J	O	U	JOURNAL	Location/Qualifiers
F	E	A	FEATURES	1..11747
S	O	U	SOURCE	/organism="Clostridium botulinum"
S	O	U	SOURCE	/strain="C-Yoichi"
S	O	U	SOURCE	/db_xref="taxon:1491"
G	E	N	gene	6..545
C	D	S	CDS	6..545
C	D	S	CDS	6..545
C	D	S	CDS	/gene="orf-22"
C	D	S	CDS	/gene="orf-22"
C	D	S	CDS	/codon_start=1
C	D	S	CDS	/transl_table=11
C	D	S	CDS	/product="ORF-22"
C	D	S	CDS	/protein_id="BAB71744.1"
C	D	S	CDS	/db_xref="GI:16580754"
C	D	S	CDS	/translation="MNDLFVAENLKHDNQNFEMSLKKRYKSKNLYYYDYN
C	D	S	CDS	KLDLVWKEILETNLFNSLEDLRKYISTIKRYCINICKKNRKDKIIYNSRVTK
C	D	S	CDS	KIDAVNVYLCVDGNFPFLDISILNYKERQIIYVMFFEGGRKDNEITPLRLSRQSIVK
C	D	S	CDS	IRISLKKLIPTVMQLVNI"
C	D	S	CDS	complement(620..2491)
C	D	S	CDS	/gene="ha-70"
C	D	S	CDS	complement(620..2491)
C	D	S	CDS	/gene="ha-70"
C	D	S	CDS	/codon_start=1
C	D	S	CDS	/transl_table=11



```

/feature="HA-70"
/protein_id="BAB71745.1"
/db_xref="GI:16580755"
/translation="MS:SIKELYTKDKSINNVNLDAGNVVNRCDGWLSPQONILG
GNISNNGCTAIYDGLRRETAAYIPTASFNERYIKNNVQVFNFTASIEPIGFE
FSKTAPEKSLIMLYQTYIIRYEIKVLQNTVERAVLVPGLGVKSEFEPSBOID
KNFTSODKILNEKFIYKIDDTIIVKESKNNNINFTSOTILPYENGLYINK
GDGMYRNDKDLIGSTISGSIQPRLRNTPLFNSTQITPQYETEARLND
AFNTQLENTTLFKVEEAPTNNKISMVNTYKELINTQNGNIDSKAAYLPSS
KCVSDAPSPQAVPTVDQDGFOTQGNENIIVGVINPSENIIEISTPIDDVYI
NIPYSIONNACYLVKNTYGVKIKTKNNLPPLIYEAGSNRNWNNISNDNIK
AIKTYLNRSDAKSYLIVLEKDKAYYIRIPOISSSTYSQLIFKRELNGISDLADST
VILNDLNTSGTHYTRQSPDGVNYSYQLTFDFNNFIASSIFSFRNNOGIGTLY
RLTISNGYLNITINNSTDLSNVNVEPISLLNGATYIFRVKVELNINNIIFDAYRNS"
complement(2505..2945)
/gene="ha-17"
complement(2505..2945)
/gene="ha-17"
/codon_start=1
/transl_table=11
/product="HA-17"
/protein_id="BAB71746.1"
/db_xref="GI:16580756"
/translation="MSSERTFLPNGYKIKLSFNSLYLITYSSGALSFNTSLDNQK
WKLXIYSSNGFRFSNVAENKYLAYNDGYFVLSSSNNSLNWPIKIAINSYIICLT
SIVNVDYANTYDNNNNITDQILLNPNEDINNQQILKLEK"
complement(2999..3856)
/gene="ha-33"
complement(2999..3856)
/gene="ha-33"
/codon_start=1
/transl_table=11
/product="HA-33"
/protein_id="BAB71747.1"
/db_xref="GI:16580757"
/translation="MSQTANDLRNNEVFFISPSNNNTNKVLDKISQSEVKLNKLJSGA
NQKWLIDYDKQKAYIKVMDNTSLIITWNPALSSVSVKTDNGDNQYVILLQNTYISR
NVIIIRNTNMLVQYKIDDTLMTYSTSSNOFFKFSNCSFKTQTSITI
KFIDKONSNNVTIWSNNGDNQKWKILYNESKMAVLTLCIKNNEVLTWFSSIGNVG
TYRTGNDQYWFYINLNDASMTYISNFSNQSKFLDVVNSGLADGTNVQVWDSNGTS
AOKWITPRL"
4119..7709
/gene="ntaha"
4119..7709
/gene="ntaha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAB71748.1"
/db_xref="GI:16580758"
/translation="MDINDDLNINSPVDKNVYIVRAKNTIFFKAKFVAPNIWAPE
RYIGEPDLIAEYKLDGDIYDSNLFQDSERENFLQAIILKLRINNTSGKQLLSLI
STALPPFYIGGGYSNPNIITFGKPKSKNKLNSVTSTIPPPGGYRETNVIESON
KNFYASNVITFGPGSINVENNYKKNDAENGCTMAEIVFQPLTYKYNKFYIDP
RMELTKLISLFLYGLIKPSDLNLYPRLTDLNKPQSQNLIDILLSSGVDFEPI
NTNPTWNTSFPNSIKMPEKNIYKTEIEGNAIGNDIKLKKOKFOINQVDLWNL
NLNFCOSFNSIPDRFSNALKHFRQYITMDYTDYNINGFVNGQINTKLELNKN
TNIISKEPVNLYNERNISLMKSNYIGDLKGTEDFYSTYKIPYNEEYFNDSO
NPLNLSISEEDSDIPBIIIDINPKYKNSDNLVFTQITSMTEVTTALSINYLQAOI
TNNEFVLSLSDPSKSVSSKSLVYSGFLDNLSYLETIKNDGPIDIDKVKYILWLKVF
KNYSFDLNLQEDISMGQINVELWFKGALNLTNSPVEEQDSGALISLKKDNL
REPNIEDDSDLSLGSFKDLANKLYEISKNIVYFKKIYIFSLDQWTEYYSQFFE
LIMAKQSILAQESLVQIKVONKFTDLSKASIPIDTLKIRETKFTIDLSNEQOIS
MNRVDNLNKAISCFVEDIYKPIFSMEKYININIKIREFTQRTNIDNEKISILI
NSYFTKIDFKFLDIOIKFNFSQVEQVMEKILSPYQLLLFASKGPNNSIEDISGK
NTLQYTESIELVGVNGESLYLKSNETIKFSNKFTNGLTNNTICFWLFTGKND
DKTLIGNKNVKNWEIYFEDNGLVPEIIDSNGOFSVYLSNIINDVNYIISVDRL
KQDILLRNKNKNAVSDIOLLSYSTNITSLYKNKNSLYVEELSNDLPHTSEBIVR
NYFSYLDNSYRDSKSLBYNNKYQLINYVPFETSLYEVNDNNKSYLSLKNYDGINI
SSVAFKLNIDESYVYKQWDECTICVDGTEKYLIDISPENRIQLVSKDNKAKKITV
NTDLFRPDCITFYSNDYKXIFSLSRDGDYNNMICNDNNKVPKGAHLWILES"
7724..11599
/gene="nt"

```

```

7724..11599
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BAB71749.1"
/db_xref="GI:16580759"
/translation="MPITINNFSYSDPVKNKNIYLDTHLNTLIANPEKAFRITGNIV
VIPDRFRNKNPNLKNPPRVTSPKSYDNPYDNTSDKDTFLKEIKLFRINSREI
GEELIYRLSDIIPPNNNTPIPTDFDVFNSVDVKTROGNNVWKTGINPSVITIG
PRENIDPENSTPKLNNTFAOEGGALSILISPREMLTYSNANDVGEGRFSKSE
FCMDPILILHESNHNAMENLYGAIENDQTSITTSNIYFSQYNKYLEAIEIYARGGP
TIDIPKASIEPEERKALDYIRSIARLNSITVFNKNIYVQNRKIYLSNVTTPVTAIINDNV
ESSEGVNKNRKFVLYNELTQITFEFNKAIYNQNNKLIYLSNVTTPVTAIINDNV
YDQNGNIPKSNLNLVFMGONLRNPALRVNPNKLYLTFKFKCAIDGRSLYKNT
LDCRELLVKNVTDFFIGDISDVKTDFLRKDNEETEVYIPDNVSDQVILSKNTSE
HGOLDLPSIDSESEILPGENOVFYDNRNTQNDVLYNSVYLESQKLSNVEDFTFTR
SIEPALDSAKVYTPPTLANKVAGVGLFILMWANDVVEDEFTNLRKDLKDISD
VSAIIPYGPALNATISVREGNEFEAPVGVTVLLEAFPEPTIPALGAFVIYSKQVE
RNEIILKIDNLCQRKWKQSDYEMWGTWLSRIITQPNNISYQWYDSLNYAGAIIKA
KIDLEYKKSQSDKENIKSQVENKLSVDKISEAMNNINKFIRECVTILFKNNLPK
VIDELNEFDRTKAKLINLIDSHIILVGEVDKAKVNNSPQNTIPFNISYTNLSL
LEDKLINEFNININDISLQNRKNTLVDTSGYNAEVSSEGDVQNPFPDFKLGSS
GEDRGKIVYNININDISVNSYESFSFIRNKVSNLPVYTIIDSVKNNSGWSIGI
ISNPLVTLKQNSDEQSINFSDISNAPGVNKFEVTVNMMGNKIYINGKLIID
TIKVKELTGTFESKTITFEINKIPDGLITSQSDNIMKATRDYIFAKELDGKDLNIL
FNSQYTNVYKDWGNDLRNKYIYVNDYLNRYMYANSROIYVFNTRNNRDNFEGY
KILIKIRGNINDIRVGGDILIFYDMTINNKAYNLFMKNETIADNHSTEDIYALGR
EQTKDINDNIIFQPMNNTIYASQIFKSNFNGENISGICISGITYRFLRGDWRHN
YLVPVTKQGNVYASLLESTHWFVFPVSE"
BASE COUNT 4574 a 1263 c 1622 g 4288 t
ORIGIN

Alignment Scores:
Pred. No.: 8,82e-152 Length: 11747
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 1 Gaps: 0

US-09-910-186a-10 (1-450) x AB061780 (1-11747)

QY 2 ThrilleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
|||||
Db 10250 ACAATACCCCTTTATATTTTTCATATACATAATAATCTTATTATAAGATATAATTAAT 10309

QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
|||||
Db 10310 GAATATTTCAATAATTAATTAATGATTCAAAAATTTTGACCTACAAAAACAGAAAAATACT 10369

QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
|||||
Db 10370 TTAGTGGATACATCAGGATATAATCCAGAACTGAGTGAAGAAGCGGATGTCAGCTTAAT 10429

QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
|||||
Db 10430 CCAATATTTTCCATTTTGACTTTAAATTAGGTAGTTACAGGGAGGATAGAGTAAAGTTATA 10489

QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
|||||
Db 10490 GTAACCCAGAAATGAATAATTTGTATATAATTTCTATGATGNAAGTTTACGATTAGTTT 10549

QY 102 TrpIleAa-gileAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
|||||
Db 10550 TGGATTAGATAATAATAATGGTAGTAATTAATTTACCTGGATATACATAATAATGATAGTGT 10609

QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
|||||
Db 10610 AAAATAACTCAGGTGGAGTATAGGTATTATTAGTAAATTTTAGTATTACTTACTTAAAA 10669

QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
|||||

```

1 (bases 1 to 12297)  
 Hauser,D., Eklund,M.W., Boquet,P. and Popoff,M.R.  
 Organization of the botulinum neurotoxin C1 gene and its associated  
 non-toxic protein genes in Clostridium botulinum C 468  
 Mol. Gen. Genet. 243 (6), 631-640 (1994)  
 94301293  
 PUBMED 8028579

2 (bases 1 to 12296)  
 Hauser,D.F. Submission  
 Direct Submission  
 Submitted (03-MAY-1993) D.F. Hauser, Inst. Pasteur, Unite des  
 Toxines Microbiennes, 28, Rue du Docteur Roux F-75724, Paris Cedex  
 15, FRANCE  
 On Jul 27, 1994 this sequence version replaced gi:509268.  
 Related sequences: X53751 & X66433.

FEATURES  
 Source  
 1. 12297  
 /organism="Clostridium botulinum phase 1C"  
 /strain="468"  
 /sub\_species="type C"  
 /db\_xref="taxon:35348"  
 /lab\_host="Clostridium botulinum"  
 289..294  
 /note="ORF22"  
 302..841  
 /note="ORF22"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA51308.1"  
 /db\_xref="GI:509269"  
 /db\_xref="SPTREMBL:Q38195"  
 /translation="MNDLSVATENLKHQNFNFTMSLKKYIEKTSKYNLYDYDYN  
 DILVHLKELIELNLKNFSELDLRKYISTSKRYCINICKKNRDKIILYNSEVYK  
 KLDVAVYSLICNFFELDLISLNYKEQIYKHFEGRKDNIAIRLRLSRQSIYK  
 IRTSLKLYPIVQVLVNI"  
 gene  
 complement(854..2787)  
 /gene="ANP-70"  
 854..906  
 /note="putative; ORF22"  
 complement(854..906)  
 /gene="ANP-70"  
 /note="putative"  
 complement(916..2787)  
 /gene="ANP-70"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA51309.1"  
 /db\_xref="GI:509270"  
 /db\_xref="SWISS-PROT:P46085"  
 /translation="MSLSIKELYTKDKSINNVLADGNYVNRGDGWLISRONQLG  
 GNSNGCTAIVGLDLRIRETATPYVYPASNEEYIKNVQNVFANFTEASEIPIGE  
 FSKTAPSKSLINFLQYTVIRVELIKVLQNTVETRAVLVPSGLGYKASIFNSEEIOID  
 KNYFISQDKLNIKRFYKKIDDTITVRESKNNNNINFTSOTILPYENGLYLVNK  
 GDGMRNTKDLTLGLLIESSTGSGIIQRLNTRTLPFNTSNPTIEFSQETTEARLND  
 AFNIQFNSTTLFKEVEEAPTKNISMKYNTVEKVELINYQNDIKDAEYLPISL  
 KGEVSDAPSPQAPVYEVTDQDGIQTGPENIIVGIPNSENIEISITPIDDTY  
 NIPTSQACVAVLFKVTNYGYTKTKNLPPLIYVEAIGSSNNRNNNSNLSNDIK  
 AIKYITGLNSDAKSVLYSLFKDNVYIRIPIQSSSTISOLIKRELGLNSLDADT  
 NVILNDLNSGTHVYRQSPQDVGNTISYQTLIPGDENNIASSIFSPTRNNQIGLTY  
 RLRESINGYNLITNNYSPLLIANNVPEISLLNGAYIFRVKVTNELNNYIIFDAYRNS"  
 complement(2801..3252)  
 /gene="ANP-17"  
 complement(2801..3241)  
 /gene="ANP-17"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA51310.1"  
 /db\_xref="GI:516172"  
 /db\_xref="SWISS-PROT:P46083"  
 /translation="MSERTFLPNGNYIKSLFSDSLYLTIVSSGALSFSSTSYSLDNQK  
 WKLEYISSNGFFSVAENPNKYLAYNDYGFYIYSSSSNSLWNPFIKAINSYIICLT  
 SINVTYDAWYIYDNNNNITDQPIILNLPFNFINNSNQLIKLEK"  
 complement(3247..3252)

terminator  
 terminator  
 CDS

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

```
gene
complement(3304. .4176)
CDS
/feature="ANTP-33"
/feature="ANTP-33"
/feature="ANTP-33"
/codon_start=1
/transl_table=1
/protein_id="CAA51311.1"
/db_xref="GI:516173"
/db_xref="SPTREMBL:Q38196"
/translation="MSQTANDLRNNEVFFISPSNNNTKVLKDISQSEVKLWKLQSGA
NOKRLAYDTNKKQYKIKVMDNTSLITWNAPLSSVSKVDTGNDQYKYLQNYLSR
NVIIRNPNPLVQYNDITLWYSTQISSNOFFKFSNCIYEALNLRNKKLQTLQNS
DRFLSKLNSOILVWLQWFOSSROKWIIEYNETKSAVTLKCOENRNYLWQNSNVY
ETYOSTSLQYWNINVLNDNDASKYILYNLDINRVLVDVNSQIANGTHVIVDSYHN
TNQOKLINLI"
complement(4171. .4176)
RBS
/feature="ANTP-33"
4313. .8017
gene
mRNA
4313. .>8017
RBS
4411. .4416
/feature="ANTP-139"
/feature="ANTP-139"
4427. .8017
CDS
/feature="ANTP-139"
/codon_start=1
/transl_table=1
/protein_id="CAA51312.1"
/db_xref="GI:516174"
/db_xref="SPTREMBL:Q38197"
/translation="MDINDLINSVDNKNVIVRARKNTTFFKAFKVPNFIWAPE
RYGEPLDIAEYKLDGSDYDNLSDQSRNENFLQAIITLLKRINTISGKQLLSLI
STAIPFYGYGSSPNIFTFGTKSNKLNLSVTSTIPFPFGYRETYIEQSN
KNFVASYVITFGGSSPNVNNVYKKNDAENGMTABIVFOPLLTYKYNKYIDP
AMELTKLILKLYLQKPSDNLVVPYRLTELDAKQFSQLANIDLLSGGVDFEFL
NLTWYFNSVFNPSIKMFKYKNVYKTELEGNALGNOLKPLKQFQINQVQDINWL
NLNFCQSFNIIIPDRSNALKHFKRYQYTYNDYTDNNGINGFVNGKINIKPLPSKN
TNISKPKVNVNENRISLMSNIYDGLKGTTFEDYSTYKIPNEYESYRFNDSQ
NPPNNISIEVDSPITDIINPKYNSDNLVFTQTSMTTEVTTHALSINYLQAI
TNNEFTLSDFSVKSKOKSLVSLFNLMSYLETIKNDGPDIDTKRYLWKLKEV
KNYFDNLTOEIDMCGINEVLFWGLKALNLTNSFVEYQDSGALSLSKKNL
REPNIIDILSDLLGLKPSDNLKLYEYKSNVYFKIYFSLFQWTEYYSOYFE
LICMAKQISLQSLVQYQVQKFTDLKASIPDOLKILRETTKFTDLSNQSIL
MNVNDFELKASICVVEDYDIPKESYKYNINIKREFIQRCTININDKESILI
NSYFTIDFLDIQSKNFFNSQVQVQKKEILSPQLLFLASKGPNINIEDLSGK
NTLYQITSELVYGVNGESLYLKSQNETIKFNKFTNGLTNFTICFWLRTGKND
DKRLIGNKVNCCWEYFEDNGLVFEIIDSNGNQSVYLSNINDNWTYISISVDRL
KQQLIFITNDKRVANSIDQILSYSTNIIISLVNKNNSIYVEELSVDNPIYSEEVIR
NFFSLDMSYTRDSKSLILEYKNLYQYNYVFPETSLYEVNDNKKSYLSLKTGDNIN
SSVKFLINIDESKVVQKWDQECICVLQDQTEKYLIDISPPNNRIQLVSSKDNKKTIV
NTDLFRPCTITFSYNDKYFSLSLRDGQENMINDNKNVFKGAHLILES"
7931. .12032
gene
mRNA
7931. .>12032
RBS
8016. .8021
/feature="BONT/C1"
8032. .11907
CDS
/feature="BONT/C1"
/codon_start=1
/transl_table=1
/protein_id="CAA51313.1"
/db_xref="GI:516175"
/db_xref="SWISS-PROT:P18640"
/translation="MPTLINNKYSQVDPVKNKILYLDTHLNTLANEPKAFRTGNW
VIPDRFSRNPKNLPRTVSPKSYDENVLSTDSKDTFLKELIKFKRLNSREI
GEELIYRUSTDIPFGNNPTINTFDFVDVNSVDKTRQGNWNVKIGSINFSVITG
PRENIDPETSTFKLTNTAAQFGALSIIISFRPLTYSNATNDVGEGRFSKE
FCMDPILILMELHNAHMYLGAIAPNDQTISSVTSNIFYSOYNVLEAYIAFGQF
TIDLTPKARVYFEKALDYRSIAKRLNSTITANPSFNKYGEYKQKILRYRFRV
ESSEGTWYRNKRYFVELYNELTOIFTEFNKAYIYNVQNRKTYLSNVYPTANTLDNV
YDIQNGFNIKPSNLNLFMSQNLSPALRKVPENKMLYLFTEFKHKAIDGSRLYNKT
```

```
LDCELLLVKNTDLPFICGDISDKVTDIFLRKDINEETEVIYVDPNVSDVILSKTSE
HGQDLDLXPSIDSESEILPGENQVFNDRNTQNDVLYNSYIYLESQLSNVEDFTFTR
SIEEALDNSAKVYIYFFTLANKVAGVGGFLKMWADVDEFTTILRRDLDKLSQ
VSAIPIYGPALNISNSVRNGNEFEAFVATGVTILLFEAFTEFTIPALGAFVIYSKVOE
RNEIITIDNLEQRIKRWKDSWMMGTWLSRIITQFNNSIYQMTDSLYNQAGALKA
KIDLEYKRYSGSKENIKSQVENKLSGVDKILSEAMNNINKFIRECSVTYLFKMLPK
VIDELNEFDNRTKAKLINLIDSHNILLGVWDKILKAKVNSFQNTIPFNIFSYNNSL
LKDIIINEYFENNINDSKILSLQNRKNTLADTSGYNAEVSSEGDYQANIPFPDFKLGSS
GEDRGKVIYQONENIYVNSYSEFSISFWIRINKWYSNLPGYIILISVYNNSSWISGI
ISNLELVITQONEDSEQISNFSDISNNAFGINKWFEFTVNNMGNMKIYINGKLID
TIKYKTEGTNFKSITITFEINKIPDTGLITSDSNINMWIRDFYIFAKELDGKDIL
FNSQYTNVNDYKWNDRYKXYVYNDIYJARYMYANSRQIVFTRRNNDNFEGY
KIIIRKRGNTDTRVGGDILYFDIMTINNKAYNLFMKNETMADHSTEDIYAGLR
BOTKDINDNIITFOIQPNNTYVYASQIFKSNFNGENISGICSTGTYRFRLGSGSWYRN
YLVPTVQGNVYASLLESTSHWGFVPVSE"
terminator
11993. .12032
/feature="BONT/C1"
BASE COUNT 4825 a 1287 c 1703 g 4482 t
ORIGIN
Alignment Scores:
Pred. No.: 9,23e-152 Length: 12297
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 7 Gaps: 0
US-09-910-186a-10 (1-450) x CBCBONT (1-12297)
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21
Db 10558 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTTATTAAGATATAATTAAT 10617
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 10618 GAATATTTCAATAATATTAATGATTCAAAATTTTGACCTACAAACAGAAAAAATACT 10677
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 10678 TTAGTGGATACATCAGGATATAATCAGAACGTGAGTGAAGAAGCGATGTTACGCTTAAT 10737
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 10738 CCAATATTTCCATTGACTTAAATTAGCTAGTCAGGGAGGATAGAGTAAGATTATA 10797
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 10798 GTAAACCCAGAAAGAAAATATGATATAATCTATGATGAAGATTTTAGCATAGTTTT 10857
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuPheGlyTyrThrIleLeuAspSerVal 121
Db 10858 TGGATTAGATAATAATAATGGTAAAGTAATTTACTGGATATATACTAATATGATAGTGT 10917
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 10918 AAAATAACTCAGGTGGAGTATAGTATATTAATAGTAATTTTATTAGTATTTTACTTAAA 10977
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
Db 10978 CAAATGAAGATAGTAGAACAAGATATAAATTTAGTATGATATATCAATAATAGTCCT 11037
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 11038 GCATACATAAATAGTGTGTGTGTAACTCTTACTACATAATATGATGGGAATATCAAGATT 11097
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 11098 TATAATAATGGAATAATTAATACATATAATAAAGTTAAGAACAATACTGGAATTAATTT 11157
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 11158 ACCAAACTACTAACATTGGAATAATAATAATCCAGATACCGGTTGATTACTTACAT 11217
```

QY 222 SerAspAsnIleAsnMetTrrPLeArgAspPheTyrIlePheAlaLysGluLeuaspGly 241  
 |||||  
 Db 11218 TCTGATACATCAATATGTCGTAAGAGATTTTATATATTTCTTAAGAATATAGATGGT 11277  
 |||||  
 QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261  
 |||||  
 Db 11278 AAGAT 11337  
 |||||  
 QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
 |||||  
 Db 11338 GGAATGATTTAAGATATATAAAGATATATATATATATATATATATATATATATAT 11397  
 |||||  
 QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 |||||  
 Db 11398 TATATGATGCGACTCAGCAAAATGCTTTTAAACACGTAGAATATATATATATAT 11457  
 |||||  
 QY 302 AsnGluGlyTyrLysIleIleIleArgGlyAsnThrAsnAspThrArgVal 321  
 |||||  
 Db 11458 AATGAAGGATATAAATATATAAAGAAATCAGAGAAATACAAATGATCTAGAGTA 11517  
 |||||  
 QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
 |||||  
 Db 11518 CGAGAGAGAT 11577  
 |||||  
 QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
 |||||  
 Db 11578 ATGAAGATGAACATATGTCAGATAATCATAGTACTGAAGATATATATATATAT 11637  
 |||||  
 QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 |||||  
 Db 11638 TTAAGAGAACAAACAAGGATATAAATGATAATATATATATATATATATATATAT 11697  
 |||||  
 QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyLysAsnIleSer 401  
 |||||  
 Db 11698 AATACCTATATATAGCATCTCAATATATTAATCAAAATTTTAAATGAGAGAAATAT 11757  
 |||||  
 QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsn 421  
 |||||  
 Db 11758 GGAATATGTTCAATAGGACTATTCGTTTAGCTTGAGCTGGAGTGATGGTATAGACA 11817  
 |||||  
 QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrThr 441  
 |||||  
 Db 11818 TATTTGGTGGCTACTGTGACGACAGAAATATATGCTTCATATAGAAATCAACATCA 11877  
 |||||  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 |||||  
 Db 11878 CATTTGGGTTTGTACCTGTAAGTGAA 11904  
 |||||  
 RESULT 8  
 LOCUS AB037920 11631 bp DNA linear BCT 24-JAN-2002  
 DEFINITION Clostridium botulinum genes for ORF-22, HA-70, HA-17, HA-33, NTNHA,  
 neurotoxin, complete cds.  
 ACCESSION AB037920  
 VERSION AB037920.1 GI:6939789  
 KEYWORDS neurotoxin; NTNHA; HA-33; HA-17; HA-70; ORF-22.  
 SOURCE Clostridium botulinum (strain:D-4947) DNA.  
 ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 REFERENCE 1  
 AUTHORS Kouguchi,H., Watanabe,T., Sagane,Y., Sunagawa,H. and Ohyama,T.  
 TITLE In vitro reconstruction of the Clostridium botulinum type D  
 progenitor toxin  
 JOURNAL J. Biol. Chem. 277 (4), 2650-2656 (2002)  
 MEDLINE 21659747  
 REFERENCE 2 (bases 1 to 11631)  
 AUTHORS Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Takizawa,J.,  
 Kawabe,T., Murakami,F., Muroga,A., Nakatsuka,M. and Ohyama,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-FEB-2000) Tohru Ohyama, Tokyo University of  
 Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido

099-2493, Japan (E-mail:t-oyama@bioindustry.nodai.ac.jp,  
 Tel:81-152-48-3838(ex.368), Fax:81-152-48-2940)  
 Location/Qualifiers  
 1. .11631  
 /organism="Clostridium botulinum"  
 /strain="D-4947"  
 /db\_xref="taxon:1491"  
 6. .545  
 /gene="orf-22"  
 6. .545  
 /genes="orf-22"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ORF-22"  
 /protein\_id="BAA90656.1"  
 /db\_xref="GI:6939790"  
 /translation="MNDLFYAIENLKHONHDFLEMSLKVKYIEKTSKKYNLYDYXN  
 DLIHMKELIIEINKNSELDLKYISTSKRYCINICKKKDKKLIYNSATYK  
 KLEAVNYSLYCEDEFELDLISILYKKEQIIYMAFFECRKNDEIARLHLRSIYK  
 IRIKSLKLYPIVMQLVNI"  
 complement(621. .2492)  
 /genes="ha-70"  
 complement(621. .2492)  
 /genes="ha-70"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-70"  
 /protein\_id="BAA90657.1"  
 /db\_xref="GI:6939791"  
 /translation="MSLSIKELYTKDKSINNVLADGNVYVNRGDGWLRSQNLG  
 GNTSNGCAIVGDLRIRETATPYYPASENEYIRNNVONVYFANFTEASEIPIGE  
 FSTASNGKLYNQYTYRVELIKVLENTVERAVLYVPSLGVAKSIENSQIID  
 KNFYSEDKCLNERFIYKKAETITAKESDNDNNTNLNTSQTILPYNGLYVINK  
 GDGYMNDKDLIGTLLEINTSGSLIQPLRNTTRPLFNTSNPLFSCETEAKLND  
 AFNIFNTSTLFEVEAPDNKNISKAYNTYKELINYNQNGNIADKAYLPULS  
 GKEVDASPAQVAVETVEQGTQGNENIIVGVNSENTEESTIPDDYTY  
 NITPSIONNACVLFVTNTGVQKINAONLPLIYIESIGSDNNIQTSLNNIK  
 AINWGTSSNAESVLYSLIKKNYVIRIPIQISSSTNQLIFKRELGNISDLANS  
 VLINLNTSGHYTROPSPDYGNVLSVOLTIPGDENNIASSIFSFRNNQIGTLY  
 RLDESINGNLIRIKNISLNNVPEISLNGATYIFRVKVTLENNYNIIFDAYNS"  
 complement(2506. .2946)  
 /genes="ha-17"  
 complement(2506. .2946)  
 /genes="ha-17"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-17"  
 /protein\_id="BAA90658.1"  
 /db\_xref="GI:6939792"  
 /translation="MSSERTFLPNNGYKIKSLFSDSLXYTSSGSLFLNTSSLDNQK  
 WKLEYISSNGPRFSNVAPNPKYAYNDYGFYLSSSNNSLWNPDKIAINSYIICTL  
 SIYVVDYAWTIYDNNNTIDPILNLPFNFINNSQILKLEK"  
 complement(3008. .3868)  
 /genes="ha-33"  
 complement(3008. .3868)  
 /genes="ha-33"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-33"  
 /protein\_id="BAA90659.1"  
 /db\_xref="GI:6939793"  
 /translation="MSQTANDLRNNEVEFFSPSNTNKVLDKISQSEVKLWKLKLSGA  
 NQWRLYTDNKQATYKVDNNTSLIITWDAPLSSVSKVTNTNTNQWILLQDYISR  
 NVILRNMPNLVQYNTDITLIVSTQTSNNQFFKSNICYEALNNRCKLQTLNS  
 DRELSKLNLSQIIVLQWDFDSSRQKWTIEYTKSAYTLKCOENRYLTWTQNSNYV  
 ETQVSDSLIQYWNINYLNDASKYILYNLQDTRVLDVYNSQTANGTHVIVDSYHGN  
 TNOQWIINLI"  
 4129. .7719  
 /genes="ntnha"  
 4129. .7719  
 /genes="ntnha"  
 /codon\_start=1  
 /transl\_table=11

QY	22	GlutyrPheAsnAsnAileAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
Db	10308	GAATATTTCAATAGTATTAAATGATTTCAAAATTTTGGCTTACAAAATAAAAAATACT	10367
QY	42	LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGluLeuAsn	61
Db	10369	TTGATGGATACATCAGGATATAACCCAGAGTACAGTAGAGCAATGTTTCAGCTTAAT	10427
QY	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle	81
Db	10428	CCAATATTTCATTTGACTTTAAATAGTATGTTTCAGGGGATGATAGAGTAAAGTTATA	10487
QY	82	ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe	101
Db	10488	GTAACCCAGAGTAAATGATGATATGATATGATGATGATGATGATGATGATGATGATG	10547
QY	102	TyrIleArgIleAsnLysTyrValSerAsnLeuProGlyTyrThrIleIleAspSerVal	121
Db	10548	TGGATTAGGATAAATAAATGGTAAATTAATTTACCTGATATACATAATGATAGTGT	10607
QY	122	LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys	141
Db	10608	AAAAATAACTCAGGTGGAGTATAGGATATTAATTAATTTTGTAGTGTACTTTAAAA	10667
QY	142	GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro	161
Db	10668	CAAAATGAAATAGTGAACAAGATATAAATTTAGTTATGATATATCAAGAATGCTGCG	10727
QY	162	GlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
Db	10728	GGATATAATAATGTTTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	10787
QY	182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe	201
Db	10788	TATATAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	10847
QY	202	SerLysThrIleThrPheGluIleAsnLysLysLeuProAspThrGlyLeuIleThrSerAsp	221
Db	10848	ACCAAACTATACATTTTCAATGAATGAATTAATTAATTAATTAATTAATTAATTAAT	10907
QY	222	SerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGly	241
Db	10908	TCTGATACATCAATATGCGATAGGATTTTATATCTTTGCTAAGAATTAGATGAT	10967
QY	242	LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp	261
Db	10968	AAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	11027
QY	262	GlyAsnAspLeuArgTyrAsnLysGlnTyrTyrMetValAsnIleAspTyrLeuAsnArg	281
Db	11028	GGAATGATTTAAGATATGATAAAGAATTAATTAATTAATTAATTAATTAATTAATTA	11087
QY	282	TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe	301
Db	11088	TATATGCTTAAAAAGGCAATGGAATTTGTTTAAATACACGTAATAATAATAATAAT	11147
QY	302	AsnGluGlyTyrLysIleIleLysArgIleGlyAsnThrAsnThrAsnThrArgVal	321
Db	11148	AATGAGGATATAAATAATTAATAAAGAATTAAGAGAAATACAAATCATACATAGATA	11207
QY	322	ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe	341
Db	11208	CGAGGAGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	11264
QY	342	MetLysAsnGluThrMetTyrAlaAspAsnHisSerThr---GluAspIleTyrAlaIle	360
Db	11265	-----GGTATGATATAAATCTTCTAGAAATCAGGACTGATTTAGTTCCTACTA	11312
QY	361	GlyLeuArgGluGlnThrLysAsp---IleAsnAspAsnIleIlePheGlnIleGlnPro	379
Db	11313	GGTGCAATCGATCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATG	11372
QY	380	MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsn	399

```

/product="NTNIA"
/db_xref="GI:6939794"
/translation="MDINDDLNINSPVDNKNVIVRAKNTNPFKAKFVAPNIWVAPE
RYGEPLDIAEYKLDGIYDSNLSODSRENFLOAIITLKLINNTISGKQLLSLI
STALPEYGVGGYSSNIFTPGKTPSKNKLNSLVTSTPPFGGYRETNWISQN
NKPYASNIIVFGSGNIVENNVCYKKADEMGTAAILFQPLPILTYNKFYIDP
AMELTKLKLISFLYIGLKYSDDLVPYRURTSLDNKKQSQLNIDLLISGGVDEFI
NTNPFWTNSFYSIKMFYKNIYETIEGNAIGNDIKLRKQKFNQSDIWNL
NLNFTSFENSIIIPDFSNALKHFRKYQTYTDGNYINGFVNGQINIKLPLSDKN
TNIISKPELVYVANNANISLMKSNYIGDLKGTIDFTSTYKIPYNEVEYFENDSD
NPLNNTSIEBVDSPILIDINPKYDSDLLFTOITSTEEVITHALPVAALQAI
ITNENFLSSDFSVKSSKSLVSPDLNLSYLETIKNDGPIDDKKYVWLKRVF
KNYSPDLINQIEDSSGNEVWFSGKALNINLNTSNFVEEYKONSGPISLSKDNL
SEPIREDDIPDSLGSFADLNKLKLYISKRVIFRKYIFNFDLQWTEYYSQFE
LIMAKOSILAESVVKQIIONKPTDLSKASIPDPLKIKETKTFIDLDNESQIS
MRVDENLAKASICVFVEDYIPKFIYSMEKYNINIKREFIORTNINDNEKSLI
NSYFKTIDFLFKLQIAKNFQSVQVMKEMLSYQILLFATRGPSNINLIEDISGK
NLTQYVESVLYVNGESLYLKSPPNETVFNFTGLTNFTICFWLFTGRDD
DKTRLGNKVNCCWEIYFEDNGLAFPEIIDSNGOBSVYLSNVNNKYIISVDRL
KQOLLIFINDKVNARVIEQLNLIYSTNVLSLVNKNKNSIYVEELSVDLPVASEVIR
NYFSLDNLRYSDKSLLEFNNKYNQINIKYFETSLIYEDNNKNSISLKNKPDGNI
PSVAFKLINIDESGYVQKDECLICVSDGTERYLIDISPENNRIQLVSKKNAKKTIV
NTDLPFRDCITFSYNDKIFSLSRDGDYNMICNDNNKVPKGAHLILKS"
/feature="nt"
/feature="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/db_xref="GI:6939795"
/translation="MTWPFQKSYSDPVDNDILYLIPONKLIITPVKAFMTNINW
VIPERFSDTNPSLSKPRPTKSYSDPSYLSLSTDEQDTFLKGIILKFLINERDI
GKKLNLVSGSPWGSPPEDPDETRHTNIAVEKPENGSKWNTNITSVLIFG
PLPNIIDTASLTQGGQSPFEFGTSLILKVAPEFLTFSQVTSNOSSAVLGKSI
FCMPDVALMHETLSHQYIGINIPSDKRIPOVSEGGFSDQGNVQFEELYFGGS
DVEIPIQERLQREKALGHYKDIKALNNIKTIPSSWSNIDKYLKIFSEKYNPK
DNTGNVYNDIKFNSLSDITNVMSVYVSOYNKRTHYFSKHYLPVFANILDDK
YTIINGNLFTKFNIGNSGNIERNPALQKLSVESVVDLFTKVCRLTRNSRDSSTC
IOVKNLTPYVADKDSIORSIFESQIITDETNVENTSDNLSLSDILDAKVPINPEAV
DPLPNVMEPLVNGEVEYDDITKVDYLSYLYLEAKLNNKNVNTITTSVEE
ALGSNKYITPLSLAKYKVGQAGFLKAWANEVDEFTTNMKKDTLSDVSBAI
IPYGPALNINSAIRNGFOAFATGAVFLLEGFFETIPALGQVTFYSSIQDERKI
IKTENCLEQVRKWKSYQWVSNWLSR:TTFRNHSIYQMSDLSYQADAKAKIDL
EYKSGSKENIKSQVENLKNLSDYKISEAMNNINKFIRECSVTYLFKMLPKYVIDE
LNKPLKTKTEILNIDSHNIIIVGEVDRLKAKVNESFENTIPFNISYTNLSLLKDM
INEXFNSINDSKILSLONKNTLMDTSYNAEYVEGNVOLNPIFPDFDKLASSGDDR
GKVIATONENTVYNNMYESPISFWIRINKVSNLPXYTIIIDSVKNNSGWSIGISNF
LVFLKONENQEQINPSYDISKNAAGYKFPVITITNMGNMNIYINGKLIDTIKY
KELTGINSKTIITFOMKNIPNTGLTSDSNINWIRDFYIFAKELDDKDNILFNSL
QTNVVKDYGNDLRYDKETMYINVMNRYMKKNGIYFNTNRKNNDNFNGYKTIIL
KRIGNTDTRVGENVLYFNFTITDNKQYSLGMYKPSRNGLDLPVLGALDDPMDEIR
KYGFTIIOCPNTEFYASQLFTLSSNATNRLGILSIGSYFSLKGLDDYWFNHEYLIPIV
KIEVASLLESTSTHWVFVASE"

```

BASE COUNT 4493 a 1254 c 1657 g 4227 t

ORIGIN

Alignment Scores:

Pred. No.:	7,46e-114	Length:	11631
Score:	1824.00	Matches:	352
Percent Similarity:	85.84%	Conservative:	36
Best Local Similarity:	77.88%	Mismatches:	56
Query Match:	76.25%	Indels:	8
DB:	1	Gaps:	4

US-09-910-186a-10 (1-450) x AB037920 (1-11631)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21

Db 10248 ACATACCCCTTTAAATATTTTTCATACATACTAATAATCTTTTATTAAGAATGATTAAT 10307

```

Db 11373 TCATACATTTTGCATATGATCAATATTTTTCAGTAAATGCAACAAT 11432
QY 400 IleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAsp---TrrTyr 418
Db 11433 AGGCTTGGAATCAATGAGTATGATATGTTTCAAACTTGAGATGACTATTGGTTT 11492
QY 419 ArgHisAsnTyrLeuValProThrValIysGlnGlyAsnTyrAlaSerLeuLeuGluSer 438
Db 11493 ATACAGAAATATTAATTCCTGTTATATAAATAGAGCATATGCTTCATATTAGAATCA 11552
QY 439 ThrSerThrHisTyrGlyPheValProValSerGlu 450
Db 11553 ACATCAACTCATGTTTGGTTTGTGACCTGCAAGTGA 11588

RESULT 9
CLOTNCO
LOCUS
DEFINITION Clostridium botulinum type D gene for neurotoxin consisting of
botulinum neurotoxin D and C1, complete cds.
ACCESSION D38442
VERSION D38442.1 GI:1374775
KEYWORDS neurotoxin consisting of botulinum neurotoxin D and C1; neurotoxin.
SOURCE Clostridium botulinum D (strain:South African) DNA.
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1
AUTHORS Morilishi,K., Koura,M., Abe,N., Fujii,N., Fujinaga,Y., Inoue,K. and
Ogumad,K.
TITLE Mosaic structures of neurotoxins produced from Clostridium
botulinum types C and D organisms
JOURNAL Biochim. Biophys. Acta 1307 (2), 123-126 (1996)
MEDLINE 96283801
REFERENCE 2 (bases 1 to 4049)
AUTHORS Morilishi,K.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1994) Kohji Morilishi, Osaka University, Research
Institute for Microbial Diseases, Research Center for Emerging
Infectious Diseases, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:Kohji@biken.osaka-u.ac.jp, Tel:81-6-6879-8343,
Fax:81-6-6879-8269)
FEATURES
Location/Qualifiers
1..4049
/organism="Clostridium botulinum D"
/strain="South African"
/db_xref="taxon:36829"
163..4020
/codon_start=1
/transl_table=1
/product="neurotoxin consisting of botulinum neurotoxin D
and C1"
/protein_id="BAA07477.1"
/db_xref="GI:1374776"
/translators="MTWVKDNYSPVNDNDILXIRIPONKLIITPPKAFMITQNIW
VIERFSDNPSPSPPTSKYQSYDPSYLSDEQKDTLKGIIKLFKRINERDI
GKLLNLYVVGSPFMDSTPFDTRHTNIAVEKFPENGSKWNTNIPSVLIFG
PLPNILYVTSGLTQGOQSNPSPGFTLSILKVAPELITPSDVTNCSAVALGKSI
FCMDPVALMHELTSLHLQYLGINIPSDRIKIPQVSEGFSDGDNVQFEELYFGGS
DVELIPQERKQREKALGHYKIDARLNKNTIPSSWSNIDKIKFISEKYNFDK
DNFGFLVNDKNSYLSOLTVMVMSYIYSSQVKNRTHYFESKHILPFAVNIIDNLI
YTINGNLTKGFNIENSGQNIERNPALQKLSSEVVDLETVCGLRLTRNSRDSSTC
IOVKNLTLPYVADKDSIQEISQITDETINVENYSDNLSDELDAKPTNPEAV
DPLLPNVNMEPLNPGVEEVYDDITKDVYDLYNSVYLYEAOKLSNNVENITLTISVEE
ALGVSNKIYTFLESLAEKVNKGVOAGLELFWNAVEVDDTTNMKMDTLDKTSDVSAI
IPVIGPALNIGNSALGNKQKQAFATAGVAFLEGGPEPTIPALGVFTYSYISQIREKI
IKTIENCLORVRKWSQDQWVSNWLSRITQFNHISTOMYDLSIQADAKRIDLI
LNKFDLKTRELNLIDSHNIIILVGEVDRKAKVNESEFENITPFNIFSYTNSLKKDM
INENFNSINDSKILSQNKNTLMDTSGYNAEVRVGNVQLNPFPPDFKLGSGDDDR
KVIKTQNEINSEYNAVWYSFSFWIRINKVSNLPGYTIIDSVKNNSGWSIGIISNF
LVFTLKNENSEODINFSYDLSKNAAGYNKPEVITTNMGNMVIYINGLIDTIKY
KELTGINSFTIIFQNKIPNGLTISDSNINMIRDFYIFAKELDDKDNILNLSL

```

```

QYTNVVKDYKGNLDLYDKYKYYMINVNMRYMKSXKNGIVENTRKNNDNFEGYKIII
KRIRGNNDTRVREGENVLFNTIDNKQISLGNIAFSPRLNGIDLLVPLGLQDPWDEIR
KYGSFTIQPCNTFDYASQLFLSSNATNRLGLISIGSYFSLGDDYWFNHEYILPVI
KIEHYASLLESTSTHWVFYPAE"
BASE COUNT 1621 a 438 c 630 g 1360 t
ORIGIN
Alignment Scores:
Pred. No.: 3,06e-114 Length: 4049
Score: 1823.00 Matches: 351
Percent Similarity: 85.84% Conservative: 37
Best Local Similarity: 77.65% Mismatches: 56
Query Match: 76.24% Indels: 8
DB: 1 Gaps: 4

US-09-910-186a-10 (1-450) x CLOTNCO (1-4049)
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2677 ACAATACCCCTTAATATTTTTCATATCTAATAAATCTTTTATAAAGATATGATTAAT 2736
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2737 GAATATTCAATAGTATTATGATTCAAAATTTTTCAGCTTACAAAATAAAAAATACT 2796
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsn 61
Db 2797 TTGATGGATACATCAGGATATAACGCAGAAAGTGAGAGTGAAGCAATGTCAGCTTAAT 2856
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2857 CCAATATTCCATTGACTTTAAATTAGGTAGTTCAGGGATGATAGAGGTAAATATA 2916
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 2917 GTAACCCAGATGAATAATATGATATATGCTATGCTATGAAAGTTTGTAGTATAGTTT 2976
QY 102 TrpIleArgIleAsnLysTyrPheValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2977 TGGATTAGGATAAATAATGGTAAAGTAATTTTACCTCGATATATATAATTGATAGTTT 3036
QY 122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 3037 AAAATAACICAGGTGGAGTATAGGTATTATTAGTAATTTTGTAGTTTACTTTAAAA 3096
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
Db 3097 CAAATGAAATAGTAGCAACAGATATAAACTTTTAGTTATGATATATCAAGAAGATGCTGG 3156
QY 162 GlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3157 GGATATATAAATGGTTTTTTTGTAACTATTACTACCAATATGATGGAAATATGATGATT 3216
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3217 TATATAATGCAAAATTAATAGATCTATATAAAGTTTAAAGAGTTAACTGGAATTAATTT 3276
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3277 AGCAAACTATAACATTTTCAATGAATAAATTCCTAAATCTGCTTAATACCTCAGAT 3336
QY 222 SerAspAsnIleAsnMetTyrPheArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3337 TCTGATAACATCAATATGTGGTAAGGGATTTTATATATCTTCTGCTAAAGAATTAGATGAT 3396
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261
Db 3397 AAAATATTATATATTATTATAGCTTGAATATACATAATGTTGTAAGAATTATTTGG 3456
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281
Db 3457 GGAATGATTTAAGATATGATAAAGATATATTACATGATTAACTGTAATATATGATAGA 3516

```

Qy	282	TyrMetTyrAlaAsnSerArgClnIleValPheAsnThrHrghArgAsnAsnAspSph	301
Dd	3517	TATATTCTTAATAAAGGAACGAAATGGTTTAAATACACTAAAATAATATGCAC	3576
Qy	302	AsnGLuGlyTyrLysLeilelelYsaArgILEargGlyAsnThrAsnAspThrArqVal	321
Dd	3577	AATGAAGGATATAAATATATATAAAGAATAGAGAAATACAATCATACTAGACTA	3636
Qy	322	ArgGlyGlyAspileuleutyPhaspMethThrlEasnAsnLysalaTyrrsnLeuPhe	341
Dd	3637	CGAGGAGAAAAATGTATTATATTTAACTACCAATTCATAACAACAATATAGTTTA	3693
Qy	342	MetLysAsnGluthMeWtyTrAlaasPasHisSerThr--GluaspIeilyralatle	360
Dd	3694	-----GGTATGTATAAACCTCTAGAAATCTAGGACTGATTTAGTCCACTA	3741
Qy	361	GlyLeuArgglucInThrLyasp---IleasnAspAsnIleilePheginlegInPro	379
Dd	3742	GGTGCAATTCGATCAACCAATGSATGAGATACGTAATATGTCGTTTTATAACAACC	3801
Qy	380	MetAsnAsnThrtYtyrTyraLSerClnIlePheLysSerAsnPheasnGlyGuasn	399
Dd	3802	TGCAATACCTTGGATTACTATGCFACAAATATTTGTCCAAGTAATGCCAACAAAT	3861
Qy	400	IleSerGlyileCyserIleelyThrtYrArgPheargLeuclgLYasp---TipTYr	418
Dd	3862	AGGCTTGAATACTATCAATTCGGTAGITATAGITTCAAATTCGGATCACATTTGGT	3921
Qy	419	ArgHisasnTyrleuValproThrvalylcngLnGlyasnTyralsarLeuLeuglus	438
Dd	3922	AATCAAGCAATATTAATCCCTGTTATAAAATAGAGCATTCCTTCATTATTAGAATCA	3981
Qy	439	ThrSerThiSTripGlypheValProValSerGlu	450
Dd	3982	ACATCAACTCATGGTCTTTGCTACCTGCAAGTGAA	4017
RESULT	10		
CLOEN			
LOCUS		4043 bp DNA linear BCT 18-DEC-2001	
DEFINITION		Clostridium botulinum type C gene for neurotoxin, complete cds.	
ACCESSION		D49440	
VERSION		D49440.1 GI:1217586	
KEYWORDS		neurotoxin.	
SOURCE		Clostridium botulinum C (strain:6813) DNA.	
ORGANISM		Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
REFERENCE	1		
AUTHORS	Moriishi,K., Koura,M., Fujii,N., Fujinaga,Y., Inoue,X., Syuto,B. and Oguma,X.		
TITLE	Molecular cloning of the gene encoding the mosaic neurotoxin, composed of parts of botulinum neurotoxin types C1 and D, and PCR detection of this gene from Clostridium botulinum type C organisms		
JOURNAL	Appl. Environ. Microbiol. 62 (2), 662-667 (1996)		
MEDLINE	96156810		
REFERENCE	2 (bases 1 to 4043) Moriishi,K.		
AUTHORS	Direct Submission		
TITLE	Submitted (23-FEB-1995) Kohji Moriishi, Osaka University, Research Institute for Microbial Diseases, Research Center for Emerging Infectious Diseases; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kohji@iken.osaka-u.ac.jp, Tel:81-6-6879-3343,		
JOURNAL	Fax:81-6-6879-3269)		
FEATURES	Location/Qualifiers		
source	1..4043 /organism='Clostridium botulinum C' /strain='6813' /db_xref='taxon:36828'		
RBS	130...136 note='SD sequence'		
CDS	146..3988 /codon_start=1 /transl_table=11		



Db	3143	ACAGGATATACAAATAAATGCGTTTTTGTGTACTATAAATAATAATAATGCGGTATATG	3202	TITLE	Organization of Gene Encoding Components of the Botulinum Progenitor Toxin in Clostridium botulinum Type C Strain 6814: Evidence of Chimeric Sequence in the Gene Encoding Each Component Published Only in Database (2000)
QY	180	LysileyrileasnglyLysleuileasphrileLysVallysgleuLeuThrglyile	199	JOURNAL	2 (bases 1 to 11614)
Db	3203	AAACTTATATAAATGGAAGAAATGAAGCAGAGTGAAGAAATGAAGATTAAATGAGGTT	3262	REFERENCE	2 (bases 1 to 11614)
QY	200	AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr	219	AUTHORS	Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Kawabe,T., Murakami,F., Nakatsuka,M. and Ohyama,T.
Db	3263	AACTTAGATAAACCACATATATTGGAATAGAGAGATAATAGAT-----	3307	TITLE	Direct Submission
QY	220	SerAspSerAspAsnIleAsnMetTrpIleArgAspPheThrIlePheAlaLysGluLeu	239	JOURNAL	Submitted (17-JAN-2000) Tohru Ohyama, Tokyo University of Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido 099-2493, Japan (E-mail:t-ohyama@bioindustry.nodai.ac.jp, Tel:81-152-48-3838(ex.365), Fax:81-152-48-2940)
Db	3308	-----GAGAATCAGATGCTTTCGATTGAGATTTTAATATTTCCTAAAGAAATTA	3358	FEATURES	Location/Qualifiers
QY	240	AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValIlyAsp	259	Source	1. .11614
Db	3359	AGCAATGACAGATATTATTTGATGAGGACAAATATTAAGAAATGTTATTAAGAT	3418	gene	/organism="Clostridium botulinum"
QY	260	TyTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu	279	CDS	/strain="C-6814"
Db	3419	TATTGGGAAATCCTTTCGATTGATACAGATATTATATTATTAATGATAATATATA	3478	gene	/db_xref="taxon:1491"
QY	280	AspArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn	299	CDS	/gene="orf-22"
Db	3479	GATAGGTATACACCTAAAGATATATACATTCATTCCTGTCAGTATCCAGATAGATCT	3538	gene	/gene="orf-22"
QY	300	AspPheAsnGluGlyTyrLysIleleleLysArgIleArgGlyAsnThrAsnAspThr	319	CDS	/codon_start=1
Db	3539	AAATTATATCTGGAATCCTATTACTATTAAATCAGTATCTGATAAGAAATCCTATAGT	3598	gene	/transl_table=1
QY	320	ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn	339	CDS	/product="orf-22"
Db	3599	AGATTATTAATGACATAATATATGTTTCATATGTTATATATAGTGGGAATATAG	3658	gene	/codon_start=1
QY	340	LeuPheMetLysGlnThrMetTyrAla-----AspAsnHisSerThrGluAsp	356	CDS	/product="orf-22"
Db	3659	ATAATAGAGATACTACATATATGCAATAGAGGAGAGAGTGTCAAAAAATGT	3718	gene	/protein_id="BAA89709.1"
QY	357	IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln	376	CDS	/db_xref="GI:6729209"
Db	3719	GTATATGCATTAATATACAGATTAATTAGTAATTTAGTATAGGT--ATATTTAGT	3775	gene	/translation="MSLSIKELYAKDKSINNVLADGNYVNRGDGWLSSRONQLG
QY	377	IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn	396	CDS	/transl_table=1
Db	3776	ATAAAAAATTTGATCTCAAAATAAATATGCTAGTCAAAATTTCTCPAGTTTATG---	3832	gene	/product="HA-17"
QY	397	GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly	415	CDS	/protein_id="BAA89710.1"
Db	3833	---AAAAATACATGCTTCTAGCAGATATATATAAACCCTTGGAGATTTCTTTT---	3883	gene	/db_xref="GI:6729210"
QY	416	AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu	435	CDS	/translation="MSSSEITPLPNGNYKIKLSFNSLYLYTSSGALSFSNTSSLDNOK
Db	3884	-----GAAAATGCATACACGCCA---GTTGCGATCACTAATATATGAGACAAA	3928	gene	/WKLEYISNNGFRFSNVAESNKLYANDYGFYILSSSSNNSLWNPKIANSYIICITL
QY	436	LeuGluSerThrSerThrHisTrpGlyPheVal 446		CDS	STVNYTDYAWTYIDNNNNITDQILNLPNFDINNSNQILKLEI"
Db	3929	CTATTATCAACTTCATCTTTTGGAAATTTATT 3961		gene	/complement(3020. .3880)
RESULT 11				CDS	/gene="ha-33"
AB037166				gene	/complement(3020. .3880)
LOCUS				CDS	/codon_start=1
DEFINITION				gene	/transl_table=1
AB037166				CDS	/product="HA-33"
ACCESSION				gene	/protein_id="BAA89711.1"
VERSION				CDS	/db_xref="GI:6729211"
KEYWORDS				gene	/translation="MSQTNANDLRNNEVFFISNSTKNKVLDKISQSEVKNSKNGLS
SOURCE				CDS	NOKWBLIYDTNKOAYKIKVMQYKISILTDAPLSSVSKYTDNTNTNNOYTWYLLDY ISR
ORGANISM				gene	/translaton="MSQTNANDLRNNEVFFISNSTKNKVLDKISQSEVKNSKNGLS
REFERENCE				CDS	1 (bases 1 to 11614)
AUTHORS				gene	Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Kawabe,T., Murakami,F., Nakatsuka,M. and Ohyama,T.



```

NVLIRNPNLVLYQWNTDITLVSTQSSNQFFKFSNCIYEALNNRNCKLQTLQNS
DRFSKLNQSIIVLWQWFSRQKWLIEYNTKSAITLKCOENRNYLTWQNSNYV
EYQTSQSLQYNNVLDNDASKYILYNLQDTRVLDVNSQIANGTNVIVDSYHGN
TNOQWIINLI"
4145..7735
/gene="ntaha"
4145..7735
/gene="ntaha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAAG9712.1"
/db_xref="GI:6729212"
/translation="MDINDDLNINSPVDNKNVIVVARKTWTFFRAFKVAPNWIATPE
RYGEPDLDAEYKLDGDIYNFLSDSRENFLOAIITLLKRINNTISGKQLLSLI
STAIPIYGVYGGYSNPSTFTFGKPKSNKLNLSVTSTIPFPFGYRETYIISPEN
NKDYAFNIVFGPNSVNNVICYKKNDAENGMTMAEILFPQLPILYKRYFIDP
AMELTKGLISLFLYGIKPSGLVYVRLRTELNDKQFSQNLIDLLISGGVDLEFI
NTNPTWNTSVFNSIKMFKYKNIYETEIEGNAJNDIKRLRQKFGNSQVDIWL
NLNYSKEPNSIIPDRESNALKHFKYKQYVMDYTDNYNINGVNGOINIKLPLSDKN
TNISKREKVNLVNNENISLMKSNYGDGLKGTEDFYSTYKIPNEEYEFYRDNDS
NPLNNISIEVDIPILIDINIKONSDNLVETQITSMTEEVYTHALPINYLOAQI
TNENFLTSSDFSQVSKSKDSLYIFSLDNLSYLETIKNDGPDIDTRKYYIWLKEVF
KNYSFDNLQEDISSGGINEVYFGLKALNINLTNSFEVEYQNSGSPSLISKKDNL
SEPIEIDIDPQSLGLSFSFDLANKLYEYISKNIYFKIYFNFLQDMWTEYISQYFE
LICMAKQSIILAEQSLKCIQNFDTLSKASIPDYLKIKETKTFDLKSKQSIS
MRYVDNLKASICVFDIYKPIYSYMEKYINNINIKURETORCTNNDKESLII
NSYFKPDLKFDLQICKNFNSQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
NTLQYTESVFLYGVNGESLYLSPNETVEFSNFTNGLNINFTICFWLAFPTKDD
KDLRLGNKNNQWEIYFDNGLVFIIDSNQNSQVSLNSVINNNWYIYISIVDRS
KDQLLITNKVNAQIEQILNIYSTNVLNSKNSIYVEELSVDYTYISVDRS
NFFSYLQNSIDRSKSLYNNKYNLYNFFPKTSLEYVDNNKNSYLSKNDGINI
PSVFKLNIIDESGYQKWDCEIICVDSGTKEYLDISENNNRIOVLVSKDNKAKTIV
NTDLFRPDCITFSYNDXVFSLSLRDGDYNNMICNDNNKPKGAHLILES"
7750..11592
/gene="nt"
7750..11592
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BAA9713.1"
/db_xref="GI:6729213"
/translation="MPITINNFNVSQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
VIPQFSDSNPNLKNPRTYSPKSGYDNYLSTSEKDTFLKEIKLFRKINSREI
GELIYRLATDIPPGNNTPINFDVDFNSVDVKTROGNNWVATGINSVITIG
RENIDTSTPKLTNTAAQGFALSIISIPREMLTYSNATNNVGEGRFSKSE
FCMDPILILHMLNHNLYGTAIPNDQRISSVTNIFYSQYKVKLEVAEYAFGGP
TIDLPSARKYFEKALDYRSIAKRLNITITANPSSFNKYIGEYKQKLIKRYEVY
ESSEVADRKFALYKELTQIFTEENYAKIYNQNRKIYLSNVITPYTANILDDNV
YDQNGFNIPKSNLNFQNGNSRNPAKRVNPNMLYLFKFKRAIDGRSLYNTK
LDRELIVKNTDLPIDIGDIDKIDINEETEVIDYDPNVSDVQVILKNTSE
HQQLDLYPILEGQSVPQNGVQYDNRQNDVYLSNYLSEKLSLSDNVEDFTFTT
SIEALDNSGVYTFYFKLADKVNTGVQGGFLMWANDVVEDEFTNILKRDITDKISD
VSATIPYGPALNISVRRNCFEAFVGTWILFAFOEFTIPALGAFVYISVKOE
RNEIYKIDNLEQRIKRWKDSYEMVGTWILRTQFNISYQMDVLSNYQADALYK
KIDLEYKISGSDENKSNQVKNLSLIDIKISEAMNNINKFIPRESYVILFKMLPK
VIDLEKPFKLTNKLINDSHNIIIVGEVDRLKAKVNFESFENTIPFNITKNSL
LKDIINEYFNKINSKLSQNKALVDYSGVNAEVRLEDQVQVNTIYTNFKLSS
GDKIIVNNNNIISAYEVSWSFWIKISDNTSHNEYTIINSKQSGWKLCIRN
GNIEWILODINRKYSLIFDYSLSHTGTNKFVFTITNNIMYKLYINGELKOS
FRIEDNLVKLDKTVFGIDENIDENOMLWDRNFISKLSNEDNIVVEGQILNRY
TKDWGNPLKEDTEYIINDYIDRYAPKSNITLVQVDPDRSKLYTNPITKSVSD
KNFYRLNGONMFMHLYNSGKTYIIRDTIYIEGRCNSKNCVYALKQSLNGNY
GIGIFSINIVSONKYSQIFSSFMKNTMLLADIYKPRFSENFENAYTPVAVTNYETKL
LSTSFWRKIFSRDFQWVE"
BASE COUNT 4499 a 1240 c 1651 g 4224 t
ORIGIN

```

Alignment Scores:  
 Pred. No.: 2,61e-50 Length: 11614  
 Score: 882.50 Matches: 198  
 Percent Similarity: 61.20% Conservative: 78

```

Best Local Similarity: 43.90% Mismatches: 148
Query Match: 36.91% Indels: 27
DB: 1 Gaps: 10
US-09-910-186A-10 (1-450) x AB037166 (1-11614)
QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21
DB 10276 ACAATACCCCTTTAATATATTTTTCATATACATAAATCTTTTATTAAAGATATAATTAAT 10335
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAspThr 41
DB 10336 GAATATTTCAATAGTATTAAATGATTCAAAATTTTGAGTTACAAAAACAAAAAATGCT 10395
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGluGluGluGluLeuAsn 61
DB 10396 TTAGTGGATACATCAGGATATAAATGCAAGAGTGGAGGTAGAGGTATGTTCAAGTAAAT 10455
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
DB 10456 ACGATATATACAAATGATTTAATTAAGTAGTTTCAGGAGAT-----AAAATTATA 10506
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
DB 10507 GTAATTTTAAATAATAATATTTTATATAGCCCTATTTATGAGAACTCTAGTGTAGTATTT 10566
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
DB 10567 TCGATTAGATATCTAAGATTTAACTAATTTCTCATATGAATATACATAAATTAATAGT 10626
QY 121 ValLysAsnAsnSerGlyTyrSerIleGlyIleLeuSerAsnPheLeuValPheThrLeu 140
DB 10627 ATAAAAACAAAATTCGGTGGAAATATATGATTAGGAATGCAATATAGAAATGCAATTTTA 10686
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
DB 10687 CAGATATTAATAGAAAATATAAAGTTTAAATTTTGATTAAGTCAATCAATAGTCAT 10746
QY 161 ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
DB 10747 ACAGGATATACAAATAAATGTTTGTCTACTATACTAATAATAATAATGCGGTATATG 10806
QY 180 LysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyLe 199
DB 10807 AAACCTTTATATAATGGAATTTAAAGCAGAGTGAAGATTTGAAGATTTAAATGAGCTT 10866
QY 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
DB 10867 AAGTTAGATAAAACCATAGTATTGGAATAGATGAGAATATAGAT----- 10911
QY 220 SerAspSerAspAsnIleAsnMetTrpIleArgPhePheIlePheAlaLysGluLeu 239
DB 10912 -----GAGAATCAGATGCTTTGGATTAGATTTAATATTTTCTPAAAGAATTA 10962
QY 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAsp 259
DB 10963 AGCAATGAGATATTAAATATGTTATGAGGACAAAATTTAAGAAATGTTATTAAAGAT 11022
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
DB 11023 TATTGGGAAATCCTTTGAAAGTTGATACAGAATATTATATATATATATGATGAAATATATA 11082
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299
DB 11083 GATAGGTATAGACCTTAAAGATATAATATCTGTTCTGTTCTAGTATCCAGATAGATCT 11142
QY 300 AspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAspThrAsnAspThr 319
DB 11143 AAATTATACTAGGAAATCCTATTACTATTAAATCAGTATCTGATAAGAAATCCTATAGT 11202
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
DB 11203 AGAATTTTAAATGGAGATATAATATGTTTCATATGTTATATATAGTGGGAAATATATG 11262

```





```
QY 102 TtpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2987 TGATTAAGATATCTAAAGATTAACTAATCTCATAAATGAATACAAATTAATTAACAGT 3046

QY 121 ValIleAsnAsnSerGlyTrpSerIleGlyIleLeileSerAsnPheLeuValPheThrLeu 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3047 ATAGAACAAATTCCTGGTGGAATATGTATAGGAATGCCAATATAGATGATTTA 3106

QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3107 CAAGATGTTAATAGAAAGATATAAAGTTTAAATTTTGAATAGTAACTTAAGTCAT 3166

QY 161 ProGlyTyr---AsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMet 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3167 ACAGATATACAAATAAAGGTTTTTGTACTATACTATATATATAATGAGGTATAG 3226

QY 180 LysIleTyrIleAsnGlyLysLeuLeuAspThrIleLysValLysGluLeuThrGlyLe 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3227 AAATCTTATATAAATGCGAATTAAGCAGAGCTCAAAAATTTGAGATTTAGATCAGGTT 3286

QY 200 AsnPheSerLysThrThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3287 AAGTTAGATAAAACCATATATTTTGGATAGATGAGATATAGAT----- 3331

QY 220 SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3332 -----GAGAATCAGATGCTTTGGATAGATGATATTTTAAAGAAATTA 3382

QY 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAsp 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3383 AGTAATGACATATTAATTTGTATATGAGGACAAATATTAAGAAATGTTATTAAGAT 3442

QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3443 TATTTGGGAATCCTTTGAAGTTGACAGAAATATTAATATTAATGATAATATATA 3502

QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3503 GATAGGTATATTCACCTCAAGATATGACTGACTGCTGCGTATCCAGATAGATCT 3562

QY 300 AspPheAsnGluGlyTyrLysIleLeileLysArgIleArgGlyAsnThrAsnAspThr 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3563 AAATATATATCTGGAATCCTATTACTATTAATCAATCAGTATCTGATAAGAACCTTATAGT 3622

QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3623 AGAATTTAAATGAGATATAATATCTCTCATATGTTATATAATAGTAGGAATATATAG 3682

QY 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis-----SerThrGluAsp 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3683 ATAATAGACATCTGATACATATATGCAACACAGGAGAGAGTGTCACAAATTTGT 3742

QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAsnIleIlePheGln 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3743 GTATATGATCAATAAATACAGAGTAATTTAGTAATTTAGTTATAGTATAGT--- 3799

QY 377 IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn 396
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3800 ATAAAAATATTTGCTCTAAAAATAAATATTTAGTCAAAATTTCT---TCTAGTTTATAG 3856

QY 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3857 ---GAAAAATACAATGCTTCTAGCAGATATATATAAACCTTGAGATTTTCT----- 3904

QY 416 AspTrpTyrArgHisAsnTyrLeuValProThrLysGlnGlyAsnTyrAlaSerLeu 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3905 -----TTTAAAAATGCATACACGCCAGTTCGAGTACT-----AATATGAACAAAA 3952

QY 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3953 CTATATCAACTTCATCTTTTGGAAATTTAT 3985
```

```
RESULT 14
LOCUS AB012112
DEFINITION Clostridium botulinum D phage gene for ORF-22, HA3, HA2, N1NH,
NTX, complete and partial sequence.
ACCESSION AB012112
VERSION 1
KEYWORDS NTX: N1NH; HA1; HA2; HA3; ORF-22.
SOURCE Clostridium botulinum D phage (strain:1873) DNA.
ORGANISM Clostridium botulinum D phage
VIRUSES 1 (sites)
NakaJima,H., Inoue,K., Ikeda,T., Fujinaga,Y., Sunagawa,H.,
Takeshi,K., Ohyama,T., Watanabe,T., Inoue,K. and Oguma,K.
Molecular composition of the 16S toxin produced by a Clostridium
botulinum type D strain, 1873
Microbiol. Immunol. 42 (9), 599-605 (1998)
AB017546
2 (bases 1 to 11584)
Inoue,K., Ohyama,T. and Ikeda,T.
Direct Submission
Submitted (11-MAR-1998) Kaoru Inoue, Okayama University Medical
School, Department of Bacteriology; 2-5-1 Shikata-cho, Okayama,
Okayama 700-8558, Japan (E-mail:osaru@med.okayama-u.ac.jp,
tel:81-86-235-7162, Fax:81-86-235-7162)
tel:81-86-235-7162, Fax:81-86-235-7162)
Location/Qualifiers
1..11584
/organism="Clostridium botulinum D phage"
/strain="1873"
/db_xref="taxon:29342"
28..567
/gene="orf-22"
28..567
/gene="orf-22"
/codon_start=1
/transl_table=11
/product="ORF-22"
/protein_id="BAA75079.1"
/db_xref="GI:4579734"
/translation="MNDLFVALENLKHNOHFNELMSLKKYIEKTSKKYNLYDYVN
DILHLMKLELLENKFNSELDLKKYISIRKYNICNCKKKRKKKRIIYNSEVYK
KLDVNVYSLKDNFELDLISILYKEKOLIIYMKFFEGKRDNEIARLRLRSQSIYK
IRIKSLKLYPIVMQLVNI"
complement(642..2513)
/gene="ha3"
complement(642..2513)
/gene="ha3"
/codon_start=1
/transl_table=11
/product="HA3"
/protein_id="BAA75080.1"
/db_xref="GI:4579734"
/translation="MSLSIKELYTKDKSINNVLADGNYVNRGDWILSRONQLG
GNISNNGCTATVGLRIRRETATPYPTASFNBEYIKNNVQNFANPTEASEIPIGE
FSKAPTSKNLYMYLOYIYIRIELIKVLQNVIVERAVLYVPSGLYKSIENFEEQID
KNFYTSQDKCILNEKFIYKIDDTITVKESKNNNINENTSQIILPENGLYLVNK
AGNYMNTKOLIGTLIESSTGSIICPRLNTRTFLNFSNPTIFSQYTTARLND
GFNIGLFTSTLFLKFEAPTNNKINIKYNTYKVELINYONGNIDDKAEYILPSL
KCEVSDAPSPQAPVPTVDQGIOTGNENIIVGVINPSENIEISTPIPDQTY
NIPTSIONACVLFKVTGKYKITTKNLPPLIYEAIGSSNRNMNLSNDNIK
AIKTYGLNRSDAKSYLVLFKDKNYRIPQISSTTSQILFKRGLNISDLAST
VNILDNATSGTHITRQSPDVGNISYQITIPGFNNIASSIFSFTRNQIGITLY
RLTESINGNITINNYSDLLNNVEPISLNGATYIFRVKVTLELNNTIIFDAYRNS"
complement(2527..2967)
/gene="ha2"
complement(2527..2967)
/gene="ha2"
/codon_start=1
/transl_table=11
/product="HA2"
/protein_id="BAA75081.1"
/db_xref="GI:4579735"
/translation="MSSERTFLPNGNKYKIKSLFSDSLYLTYSGLSGLSPNTSSLDNOK
```



[illegible]

```

/db_xref="GI:15982937"
/translation="MPVTINNEVNDPIDNNILMPPFPARGMRYKAFKTDIRIM
IIPERYFGKPEDFKNSSINRDVCEYDYLNDKKNIELOQMILKFRKRSK
IPGKPLEMNETIDIGIQRHFASRFGFGIMQMKPEYVSFVFNVENGASIFNR
FGPGLVILMHLEIHLVHGLGVKLVNDLPIVNEKPFQMOSTDAIOABELYFGGD
GYSDPALDKSIYDKVLQNGRGIQVKNLVCISDPNININIKNFKDKIKFEVDS
PSIISPTSDKSIYDKVLQNGRGIQVKNLVCISDPNININIKNFKDKIKFEVDS
EGKYSIDVSEFDKLYSKLMGFTENIAENYKIKTRASYPSDSLPPVKIKLNDIET
TIEBGFNISDNKMEKTRGONKAINKOAYEISEKHLAYKIQMCKSVKAPGICLDV
NEDUFFIADKNSFSDLSKNERIAYNTQNYIENDFSINELIDLDLISKELPSEN
KSLYDFVPMYVKKQQA.KKIFTDENTIFYLYSOTFPLDIRDISUTSSFDLALFS
NKVVSFMSYD IATKNAVEAGLQVNDFVBEANKSSTMKIAOTISLIVPIY
GLALKVNEAKGNFNEAELAGASTILIEFPELLIPVVGAFLLSEIYDNKKKIIETI
NSALTIKDEKWIIDMGLIYVAOMSLTVENTQYTIKEGMKALNYOQALEELIKYKNI
YSEKERSNINIDVNSKLNREGINQADININNEFNECSVSYLMKKMPLIAVEKLDF
DNTURKMLNIDENKYLIGAEYERKSKYKYLATISYFPLSYTYNTTILIEFKFY
NSDILNINILNRYRUNKLIDUSGAKVEYDGVKLNKOFKLTISANSKIRVON
QNIIFNSMFDQFVSWEIPIRYKNDGTONYIHNEYITINCMKNNSGWIKISIRGNMII
WTLIDNGKIKSVFFYSIKEDISYINRWFTYITNNSNAKVIYINGKLESIHDIRI
IREVIANDEIIFKLDINDRTOFIWMKVSFINTELQSOTNEEYIKIYSEYKLDWF
GNPLWKNKEYEYINMAGNKSYIKLKKDSVGLBITRKSYNQNSKYINRLOYIGERFI
IRKKSQSINDIDIVKEDQYIYLDIFNLNQRWYMYKFKKEEKLFLAPISDSDEF
YNTIQIKEYDEQTYSCQLLFFKKDBESDEIGLVIHRRFYESGVFVEYKDYFCISK
YLKEVKKRPKNLGNQWQFIPKDBGWTI"
BASE COUNT      1611 a      367 c      616 g      1282 t
ORIGIN

Alignment Scores:
Prd# No.:      Length:      3876
Score:         610.00      Matches:      168
Percent Similarity:      50.81%      Conservative:      84
Best Local Similarity:    33.87%      Mismatches:     134
Query Match:      25.51%      Indels:         90
DB:              1          Gaps:         20

US-09-910-186A-10 (1-450) x AF300465 (1-3876)

QY  2  ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db  2503 AGTATACCAATTTGATCTTTCACACTACTACTAAATAACATACATACATAAGATATTAAAT 2562
QY  22  GluTyPheAsnAsnIleAsnAspSerTysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db  2563 AAATAATAAGCGATATTTTAAATAATATATCTTAAATTTAAGATATAGGATAATAAG 2622
QY  42  LeuValAspThrSerGlyTyrAsnAlaGluValSerGluClyAspValGlnLeuAsn 61
Db  2623 TTAATAAGATTTTATCAGGATATGGCGAAAGGTAGAGGTATATGATGGGTCAAGCTTAAT 2682
QY  62  ProIlePheProPheAspPheLysLeuGlyLysSerSerGlyGluAspArgGlyLysValIle 81
Db  2683 GAT-----AAAATCAATTTAAATTAACATAGTTCAGCA-----AATAGTAAGATTAGA 2730
QY  82  ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db  2731 GTGACTCAAAATCAGAATATCATATTTAAATAGTAGTGTCCTTGATTTAGCGGTATGTTT 2790
QY  102 TrpIleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyr 115
Db  2791 TGGATTAAGATACCTTAATATAAGAAATGATGGTATACAAATATATATTCATAATGAATAT 2850
QY  116 ThrIleIleAspSerValIysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhe 135
Db  2851 ACAATAATTAATTTGATGAAATAAATTCGTGATGGAAATAATCTATTAGGGGTATATG 2910
QY  136 LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleIleAsnPheSerTyrAsp 155
Db  2911 ATAAATGATGACCTTTAATTTGATATAAATGGAATAAATCAATCAGTATTTTTCGAATATAGC 2970
QY  156 IleSerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsn 174
Db  2971 ATAAAAAGAGATATATCAGAGATATAAATAGATGGTGTTCCTTGTAACTACTTACTAAATAT 3030

```

QY 175 MetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLys 194  
Db 3031 ---TCGGATAACGCTAAATTTATATTAATGTAAGTAGAATCACATATAGATATTAGA 3087  
QY 195 GluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAsp 214  
Db 3088 GATATAAGAGAAGTTATGCTTAATGATCAATAATATTAAATTAGAT----- 3135  
QY 215 ThrGlyLeuIleThrSerAspSerAspAsnIleAsn-----MetTrpIleArg 230  
Db 3136 -----GGTAATATAGATAGAACACACAGATTCATTGGATGAAA 3171  
QY 231 AspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSer 250  
Db 3172 TATTCAGTATTTTAAACGGAATTAAGTCAATCAATATTAAGAAATATATAAAATT 3231  
QY 251 LeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGlu 270  
Db 3232 CAATCATATAGCGAATATTAAAGATTTTGGGGAATCCTTTAATGTACATAAAGAA 3291  
QY 271 TyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetIyr-----AlaAsn 286  
Db 3292 TATATATGTTTAAATCGCGGGAATAAATATCATATATTAACCTAAAGAAAGATTCATCT 3351  
QY 287 SerArgGlnIleValPheAsnThrArgArgAsnAsn----- 299  
Db 3352 GTAGGTGAATTTTACACCGTAGCAATATATCAAAATTCCTCAATATATAAATTAGA 3411  
QY 300 AspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThr----- 316  
Db 3412 GATTTATATATTGGAGAAAATTTATTATAAGAGA---AAGTCAAAATTCCTCAATCTATA 3468  
QY 317 AsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLys 336  
Db 3469 AATGATCATATAGTTAGAAAGCAAGATTATATATCTAGATTTTATTATTAATCAACAA 3528  
QY 337 AlaTyrAsnLeuPheMetLysAsnGlnThrMetTyrAlaAspAsnHisSerThrGluAsp 356  
Db 3529 GAGTCGAGAGATATATG----- 3546  
QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLys-----AspIleAsnAspAsn 372  
Db 3547 ---TATAATATTTTAAAGAGAGAGAGAAAATGTTTTAGCTCCTATAGTGATCT 3603  
QY 373 IleIlePhe-----GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSer 388  
Db 3604 GATGAGTTTACATACTATACATAAATAAGATATGATGAACACCAACATATAGTTGT 3663  
QY 389 Gln---IlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly 407  
Db 3664 CAGTTCCTTTTAAAAA-----GATGAAGAAAGTACTGATGATAGGATGATTGTT 3717  
QY 408 ThrTyrArgPheArgLeuGlyGly----- 415  
Db 3718 ATTCATCGTTTCTACCACTCGAATTCGTATTAAAGAGATATAAGATTATTTTGTATA 3777  
QY 416 ---AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSer 434  
Db 3778 AGTAAATGGTAC-----TAAAGAGGTAAAAAGGAACCAACCATATAATICA 3822  
QY 435 LeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450  
Db 3823 AAATG-----GGATGTAATTGGCAGTTTATCTCTAAAGATGAA 3861

Search completed: November 7, 2002, 18:38:07  
Job time : 2772 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 16:30:36 ; Search time 253 Seconds  
(without alignments)  
4005.531 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391

Sequence: 1 MTIPNIFSVTNSLLKDI.....NYASLLESTHMGFVPSV 450

Scoring table:

BLOSUM62	
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Fgapop 6.0, Fgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlin  
-Q=/cgn2\_1/USPTO\_spool/US09910186/runat\_04112002\_111616\_20826/app\_query.fasta\_1.647  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NXPEN=2000000000  
-USER=US09910186 -CGCN\_1.1.125 -runat\_04112002\_111616\_20826 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -NAIT -LONGLOG -DEV\_TTIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*

1:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2391	100.0	1371	21	AAZ87214	DNA encoding synth
2	2391	100.0	1371	22	AAA54486	Botulinum toxin hea
3	2386	99.8	3950	20	AAZ25521	DNA coding for mod
4	2319	97.0	1502	19	AAV30588	Clostridium botuli
5	1589	66.5	1208	21	AAZ98630	C. botulinum type
6	1589	66.5	1440	21	AAZ98631	Type D BoNT non-to
7	868.5	36.3	1374	22	AAA54487	Botulinum toxin hea
8	803.5	33.6	1469	19	AAV30591	Clostridium botuli
9	589.5	24.7	1400	21	AAZ87215	DNA encoding synth
10	589.5	24.7	1400	22	AAA54488	Botulinum toxin hea
11	586.5	24.5	2532	21	AAZ87218	DNA encoding nativ
12	586.5	24.5	3891	17	AAZ29244	C. botulinum type
13	586.5	24.5	4835	21	AAZ64582	BoNT/A neurotoxin
14	583	24.4	4017	22	AAZ58862	C. botulinum BoNT/E
15	581	24.3	3876	19	AAV30579	Clostridium botuli
16	578.5	24.2	1472	19	AAV30585	Clostridium botuli
17	576	24.1	1463	19	AAV30584	Clostridium botuli
18	575.5	24.1	1347	21	AAZ87217	DNA encoding synth
19	575.5	24.1	1368	22	AAZ54491	Botulinum toxin hea
20	575.5	24.1	1535	19	AAV30586	Clostridium botuli
21	565	23.6	1526	19	AAV30580	Clostridium botuli
22	558.5	23.4	1402	17	AAZ29246	Type A neurotoxin
23	558.5	23.4	1402	19	AAV30572	Clostridium botuli
24	558.5	23.4	1546	19	AAV30575	Clostridium botuli
25	557.5	23.3	1317	21	AAZ87220	DNA encoding nativ
26	557.5	23.3	1330	17	AAZ29245	Type A neurotoxin
27	557.5	23.3	1330	19	AAV30571	Clostridium botuli
28	557.5	23.3	1332	22	AAA54482	Botulinum toxin hea
29	557.5	23.3	1338	21	AAZ87216	DNA encoding synth
30	557.5	23.3	1351	19	AAZ30576	Clostridium botuli
31	556.5	23.3	1323	22	AAA54483	Botulinum toxin hea
32	556.5	23.3	1326	22	AAA54484	Botulinum toxin hea
33	553.5	23.1	1278	22	AAZ54489	Botulinum toxin hea
34	550	23.0	1547	19	AAV30581	Clostridium botuli
35	546	22.8	1293	18	AAZ48100	Immunogenic type F
36	546	22.8	1313	18	AAZ48101	Immunogenic type F
37	546	22.8	1314	22	AAA54499	Botulinum toxin hea
38	546	22.8	1317	22	AAA54490	Botulinum toxin hea
39	542	22.7	1341	21	AAZ87213	DNA encoding synth
40	542	22.7	1341	22	AAA54485	Botulinum toxin hea
41	539.5	22.6	3509	19	AAZ26289	Recombinant botuli
42	531	22.2	1317	21	AAZ87216	DNA encoding synth
43	511.5	21.4	1460	19	AAV30593	Clostridium botuli
44	492	20.6	1347	22	AAA54589	Sequence encoding
45	471	19.7	1332	22	AAA54588	Sequence encoding

# ALIGNMENTS

RESULT 1	
AAZ87214	
ID	AAZ87214 standard; DNA; 1371 BP.
XX	XX
AC	AAZ87214;
XX	AC
DT	08-MAY-2000 (first entry)
XX	XX
DE	DNA encoding synthetic BoNT serotype C (BoNTC) Hc fragment.
XX	XX
KW	Botulinum neurotoxin; heavy chain; BoNT; serotype C;
KW	C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW	VEE; botulinism; vaccine; diagnosis; drug screening; ds.
XX	XX
OS	Clostridium botulinum.
XX	OS
XX	Synthetic.
FH	Key
	Location/Qualifiers



FT CDS 10..1362  
 FT /\*tag= a  
 FT /product= "Synthetic botulinum neurotoxin serotype C  
 FT (BoNTC) heavy chain C-terminal fragment (Hc)"  
 XX WO200002524-A2.  
 PN 20-JAN-2000.  
 PD 09-JUL-1999; 99WO-US15570.  
 PF 10-JUL-1998; 98US-0092416.  
 PR 12-MAY-1999; 99US-0133870.  
 XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 PA  
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh WT, Smith L;  
 XX WPI; 2000-160827/14.  
 DR P-PSDB; AAY77136.  
 XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 PT toxin serotypes A-G, is used for inducing an immune response against  
 PT botulinum .  
 XX Disclosure; Page 41-42; 54pp; English.  
 PS  
 CC The invention relates to novel vaccines that induce a protective immune  
 CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
 CC DNA construct comprising a vector, and at least one nucleic acid  
 CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 CC serotypes A-G. In preferred embodiments of the invention, the vector is  
 CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
 CC this vector results in the production of large amounts of a protein  
 CC encoded by a sequence cloned into the replicon. The constructs are used  
 CC to produce vaccines against botulism. The proteins can also be used as  
 CC diagnostic tools for the diagnosis of botulism. The transformed host  
 CC cells can be used to analyse the effectiveness of drugs and agents which  
 CC inhibit toxin effects. The vaccine currently used against botulism is  
 CC dangerous and expensive to produce, and contains formalin, which is very  
 CC painful for the recipient. Also, the vaccine is incomplete, in that only  
 CC 5 of the 7 serotypes are represented in the formulation. The novel  
 CC vaccine of overcomes these problems, as it is easily purified, and  
 CC available in large quantities. It is also expressed in the lymph nodes  
 CC for a better immune response. Sequences AAZ87212-287217 represent  
 CC synthetic DNA sequences encoding BoNT Hc fragments used in the present  
 CC invention. These were optimised for codon usage for expression in yeast.  
 XX  
 SQ Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;

## Alignment Scores:

Pred. No.: 1,42e-205 Length: 1371  
 Score: 2391.00 Matches: 450  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-910-186a-10 (1-450) x AAZ87214 (1-1371)

QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20  
 Db 10 ATGACCATCCCATTCACATCTCTCTACACCAACACTCCCTGTTGAAGGACATCATC 69  
 QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40  
 Db 70 AACGAGTACTTCAACAACATCAAGACTCCCAAGATCTGTCCTGCGAGACCCGTAAAGAC 129  
 QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60  
 Db 130 ACCTTGGTCGACACCTCCGGTTACACGCCGAGGCTCCCGAGGGGCTGACGTCACGCTG 189

QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80  
 Db 190 AACCAATCTCCCATTCGACTTCAAGCTGGGTCTCCGGTGAGGACAGAGTAAAGTTC 249  
 QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100  
 Db 250 ATCGTCACCCAGAACGAGAACATCGTCTACAACTCCATGTACGAGTCCCTCTCCATCTCC 309  
 QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120  
 Db 310 TTCGTGATCAGAAATCAACAAGTGGTCTCCAACTTCCAGGTTACACCATCATCGACTCC 369  
 QY 121 ValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140  
 Db 370 GTCAGAACAACTCCGGTGTGTCATCGGTATCATCTCCAACCTCTCTGCTCTTCCACCTG 429  
 QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAla 160  
 Db 430 AAGCAGAACGAGGACTCCGAGCAGTCCATCACTTCTCTACGACATCTCCAACAACGCT 489  
 QY 161 ProGlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180  
 Db 490 CCTGGTTACAAAGTGGTCTTCGTCCACCGTCAACAACATGATGGTAACATGAAG 549  
 QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200  
 Db 550 ATCTACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAAC 609  
 QY 201 PheSerLysThrIleThrPheGluIleAspLysIleProAspThrGlyLeuIleThrSer 220  
 Db 610 TTCCTCCAGACCATCCTTCGAGATCAACAAGTCCAGACACCGGTCTGTACCTCTCC 669  
 QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240  
 Db 670 GACTCCGCAACATCAACATGTGATCGGTGACTTCTACATCTTCGCCAAGGAGTTGGAC 729  
 QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260  
 Db 730 GGTAAAGACATCAACATCCCTGTTCAACTCTTCGAGTACACCAACCTGCTCAAGGACTAC 789  
 QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGlyTyrTyrMetValAsnIleAspTyrLeuAsn 280  
 Db 790 TGGGTTAAGCACTTCGAGATACACAGAGGAGTACTATGTTCAACATCGACTACTTGAAC 849  
 QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAsp 300  
 Db 850 AGATACATGTACGCCAACTCCAGACATCGTCTTCAACACCAAGAGTACACACACGAC 909  
 QY 301 PheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320  
 Db 910 TTCACGAGGGTTACAAGATCATCATCAAGGTATCAGAGGTAACACCAACGACACACAGA 969  
 QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
 Db 970 GTCAGAGGTGTGATCATCCCTGTTACTTCGACATGACTATCAACAACAGGCTACACCTG 1029  
 QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
 Db 1030 TTCATGAAGAAGACGACCATGTACGCCGACACCACTCCACCGGAGGACATCTACGCCATC 1089  
 QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380  
 Db 1090 GGTCTGCTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCAAAG 1149  
 QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
 Db 1150 AACACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACITCAACGGTGAGACATC 1209  
 QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHis 420  
 Db 1210 TCCGSTATCTCTTCATCGGTACCTACAGATTCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTG 1269  
 QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440

```
Db 1270 AACTACTTGGTCCCAACGTCAAGCAGGGTAACACTGCTCTTGGTGGAGTCCACTTCC 1329
|||||
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
|||||
Db 1330 ACCCACTGGGGATTGTCGCCAGTCCCGAG 1359
RESULT 2
AAA54486
ID AAA54486 standard; DNA; 1371 BP.
XX AC
XX AAA54486;
XX 11-APR-2001 (first entry)
DE Botulinum toxin heavy chain C-terminal coding sequence (serotype C).
XX Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX Synthetic.
OS Clostridium botulinum.
XX FH Key
XX Location/Qualifiers
FT 10..1362
FT CDS
FT /*tag= a
FT /product= H_C peptide fragment
XX WO200067700-A2.
XX 16-NOV-2000.
XX 12-MAY-2000; 2000WO-US12890.
XX 12-MAY-1999; 99US-0133865.
XX 12-MAY-1999; 99US-0133865.
XX 12-MAY-1999; 99US-0133867.
XX 12-MAY-1999; 99US-0133868.
XX 12-MAY-1999; 99US-0133869.
XX 12-MAY-1999; 99US-0133873.
XX 29-JUL-1999; 99US-0146192.
XX PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX P-PSDB; AAB04092.
XX WPI; 2001-016048/02.
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX Claim 2; Fig 5a; 73pp; English.
XX Botulin neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
```

```
SQ Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;
Alignment Scores:
Pred. No.: 1.42e-205 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-910-186A-10 (1-450) x AAA54486 (1-1371)
QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIle 20
Db 10 ARGACATCCCATTCACATCTTCTCCTACCAACCACTCCCTGTTGAAGGACATATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTAGTCTCAACACATCAAGACATCCCAAGATCTGTCCTCCAGAACCGGTAAGAAC 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTCGACACCTCCGGTTACACGCCGAGGTCTCCGAGGAGGTGACGTCACGTC 189
QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCGCGTGGAGGACAGGTAAGTTC 249
QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Db 250 ATCGTCACCCAGACGAGACATCGTCTACCACTCCATGACGAGTCTCTCCATCTCC 309
QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
Db 310 TTCGTGATCAGAATCAACAAGTGGGTCTCCAATTCGAGGTACACCATCATCGACTCC 369
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGAACAACTCCGTTGGTTCATCGGTATCATCTCCAACTTCTCTGGTCTTCAACCTG 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 430 AAGCAGAACGAGGACTCCGACGACGTCATCAACTCTCCACGACATCTCCACACAGCT 489
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CCTGGTTACAAAGTGGTCTTCGTACCGTCACCAACACATGATGGTAAACATGAAG 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Db 550 ATCTACATCAACGGTAACTGATCGACACCATCAAGTCAAGGAGGTGACCGGTATCAAC 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCAAAGACCATCACCTTCGAGATCAACAAGATCCGACACCGGTCTGATCACTCC 669
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCCGACACATCAACATGTGGATCGGTCACTCTACATCTTCGCCCAAGGAGTGGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyr 260
Db 730 GGTAAGGACATCAACATCTCTTCAACTCTTGCAGTACACCAACGTCGTCAGGACTAC 789
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
Db 790 TGGGTACGACCTGAGATACAAAGGAGTACTACATGGTCAACATCGCTACTTGAAC 849
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTAGCCCACTCCAGACAGATCTCTTCAACACGACGCTAAACAACACGAC 909
QY 301 PheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
```

```

|||||
Db 910 TTCACGAGGTTACAGATCATCATCATACCGTATCAGAGGTACACCAACGACACACAGA 969
Qy 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Db 970 CTCAGAGTGTCATCTCTACTTCGACATGACTATCAACAACAAGCCCTACAACCTG 1029
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1030 TTCATGAGACGAGACATCTAGCCGCGCAACCACTCCACCGAGGACATCTACGCCATC 1089
Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet 380
Db 1090 GGTCTGGTGGACGAGACCAAGGACATCAAGCACAACATCATCTCCAGATCCAGGCAATG 1149
Qy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1150 AACACACTTACTACTAGCTTCCAGATCTTCAGTCCCACTTCACCGGTGAGACATC 1209
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHis 420
Db 1210 TCCGTATCTCTCCATCGGTACCTACAGATTCGCTGGTGGTGCATGCTGTACAGACAC 1269
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1270 AACACTTGGTTCCACTGTCAGCAGGGAACCTACGCCCTCTGCTGGAGTCCACTTCC 1329
Qy 441 ThrIleTrpGlyPheValProValSerGlu 450
Db 1330 ACCCACTGGGATTCGTCGCCAGTCTCCGAG 1359

RESULT 3
AAX25521
ID AAX25521 standard; DNA; 3950 BP.
AC AAX25521;
XX
XX 02-AUG-1999 (first entry)
XX DNA coding for modified botulinum toxin rBoNT/C.
XX Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery;
KW mutant; ss.
XX
XX Clostridium botulinum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 32..3907
FT /*tag= a
XX
XX WO9920306-A1.
XX
XX 29-APR-1999.
XX
XX 16-OCT-1998; 98WO-0521897.
XX
XX 20-OCT-1997; 97US-0954302.
XX
XX (UWJE-) UNIV JEFFERSON THOMAS.
XX
XX Kiyatkin N, Maksymowych A, Simpson L;
XX PI
XX WPI; 1999-302646/25.
XX
XX P-PSDB; AAY05814.
XX
XX Modified toxin useful for systemic delivery of oral vaccines and
XX therapeutic agents
XX
XX Example 1; Page 26-31; 37pp; English.
XX
XX This DNA sequence codes for a modified serotype C botulinum toxin,
XX termed rBoNT/C (see AAY05814), in which amino acids His-229, Glu-230

```

```

CC and His-233 of the native sequence are substituted by Gly, Thr and
CC Asn, respectively, i.e. the zinc binding motif (see AAY05814) of
CC the light chain holotoxin is modified, resulting in loss of
CC endoprotease activity. DNA coding for the modified botulinum toxin
CC was assembled from 3 separate toxin fragments using PCR and
CC site-directed mutagenesis. The modified recombinant botulinum
CC toxin maintains its ability to translocate from the gut into the
CC general circulation but is non-toxic. It can be used as an oral
CC vaccine for antigenic peptides including botulinum toxin (i.e. an
CC oral vaccine for botulism) or for the oral delivery of other
CC therapeutic agents to the general circulation.
XX
XX Sequence 3950 BP; 1589 A; 410 C; 616 G; 1335 T; 0 other;
SQ

```

## Alignment Scores:

Pred. No.:	1-43e-204	Length:	3950
Score:	2386.00	Matches:	449
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.79%	Indels:	0
DB:	20	Gaps:	0

US-09-910-186a-10 (1-450) x AAX25521 (1-3950)

Qy	2	ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn	21
Db	2558	ACAATACCCCTTAAATATTTTTCATATATCTTAATCTTTTAAAGATATAATTAAT	2617
Qy	22	GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
Db	2618	GAATATTTCAATATATATGATTCAAAATTTTCAGCTACAAACAGAAAAATACT	2677
Qy	42	LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn	61
Db	2678	TTAGTGATACATCAGGATATAATGCAGAGTGCAGTGAAGAAGGCGATGTCAGCTTAAT	2737
Qy	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle	81
Db	2738	CCAATATTTCCATTTGACTTTAAATAGGTAGTTCAGGAGGAGTAGAGTAAGTTATA	2797
Qy	82	ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe	101
Db	2798	GTAACCCGAATGAAAATATTTGTATATATCTATCTATGATAAGATTTTACGATTAAT	2857
Qy	102	TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal	121
Db	2858	TGGATTAGAATAAATAAATGGTGAAGTAATTTACCTGGATATATCTATAATTCATGTT	2917
Qy	122	LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys	141
Db	2918	AAAAATAACTCAGGTGAGTATAGGTATTTAGTAAATTTTCTAGTATTTACTTTAAAA	2977
Qy	142	GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro	161
Db	2978	CAAAATGAAGATAGTGAACAAAGTATAAATTTTAGTATATGATAATCAAAATATGCTCT	3037
Qy	162	GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
Db	3038	GGATACATAAATGGTTTTTTTGTAACTGTACTAACAAATATGATGGAAATATGAAGATT	3097
Qy	182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe	201
Db	3098	TATATAATGGAAAATTAATAGATCTATATAAAGTTTAAAGAACTAACTGGAATTAATTT	3157
Qy	202	SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp	221
Db	3158	AGCAAACTATAACATTTGAAATAAATAAATTCAGATACCGGTTTGATTACTTCAGAT	3217
Qy	222	SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly	241
Db	3218	TCTGATAACATCAATATGTCGATAAGAGATTTTATATATATTTGCTAAAGAATTAGATGT	3277
Qy	242	LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp	261

```

Db 3278 AAAGATATTAAATATATTATTAATAGCTGCAATATATACTAATGTTGTAAGAGATTATGG 3337
QY 262 GYASnAspLeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeuAsnArg 281
Db 3338 GGAATGATTTAAGATATAAATAAGAAATATTATATGGTTAATAGATTATTTAAATAGA 3397
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsgAsnAsnAspPhe 301
Db 3398 TATATGATGCGAAGTCAGCAAAATTCCTTTTAAATACACGTAGAAATATAATGACTTC 3457
QY 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 3458 ATGAAGGATATAAATATATAAATAAGAAATCAGAGGAAATCAAAATGATAGTAGTA 3517
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3518 CGAGGAGGAGATATTATATTTTATGATGACATTAATAACAAAGCATATAATTTGTTT 3577
QY 342 MetLysAsnGluThrMetTyrAlaAsnAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3578 ATGAAGAAATGAACATATGTCAGATATCATAGTACTGAAGATATATATGCTATAGGT 3637
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnProMetAsn 381
Db 3638 TTAAGAGAACAAACAAGGATATAATCATATATATATTTCAATACACCAATGAAT 3697
QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 3698 AATACTATATATATAGCATCTCAAAATATTAATCAAAATTTTAAATGGAGAAATATTCT 3757
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421
Db 3758 GGAATATGTTCAATAGGTACTTATCGTTTATAGACTTGGAGGTGTTGTTATAGACAAAT 3817
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3818 TATTTGGTGCCTACTGTGAAGCAAGAAATATGCTTCATTATTAGAATCAACATCAACT 3877
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 3878 CATGGGGTTTTGTACCTGTAAGTAA 3904
RESULT 4
AAV30588
ID AAV30588 standard; DNA; 1502 BP.
XX
XX AAV30588;
AC
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type C1 toxin C fragment gene in pETHisb.
XX
KW Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW botulism; Botc; ds.
XX
OS Clostridium botulinum serotype C1 Stockholm strain.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH 108..1463
FT CDS /*tag= a
FT
FT
FT
PN W09808540-A1.
XX
XX 05-MAR-1998.
PD
XX
XX 28-AUG-1997; 97WO-US15394.
XX
XX 28-AUG-1996; 96US-0704159.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
PA
XX

```

```

PI Thalley BS, Williams JA;
XX
DR WPI: 1998-230234/20.
DR P-PSDB; AAW68397.
XX
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 46; Page 339-341; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype C1
CC (Stockholm strain) neurotoxin fragment C gene contained in plasmid
CC pETHisb. The encoded BotC fragment C polypeptide (see AAW68397) has a
CC His-tagged N-terminal extension. The vector was used to express
CC native (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins, especially type B and type E toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred
CC hosts for production of recombinant proteins are E. coli, insect
CC cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are
CC useful in the treatment of humans and animals at risk of
CC intoxication with clostridial toxin.
XX
SQ Sequence 1502 BP; 600 A; 155 C; 249 G; 498 T; 0 other;
Alignment Scores:
Pred. No.: 4.67e-199 Length: 1502
Score: 2319.00 Matches: 439
Percent Similarity: 98.66% Conservative: 3
Best Local Similarity: 97.99% Mismatches: 5
Query Match: 96.99% Indels: 1
DB: 19 Gaps: 0
US-09-910-186A-10 (1-450) x AAV30588 (1-1502)
QY 4 PropHeAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleLeasnGluTy 23
Db 152 CCATATCGAAGTCGTCATATGGCTAGCTAGCTTTATTAAGATATATAATTAATGATA 211
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 212 TTTCAATATATTAATGATTCAAAAATTTTGAGCCTCAAAACACAGAAAAATACTTTAGT 271
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProil 63
Db 272 GGATACATCAGATATATATGAGAGTGAGTGAGAGGCGATGTTGAGCTTAATCCAAAT 331
QY 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh 83
Db 332 ATTTCATTTGACTTTAAATTAGGTAGTTTCAGGGGAGGATAGAGGTAAAGTTATAGTAAC 391
QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTyrPil 103
Db 392 CCAGATGAATAATATGATATATATCTATGATGATCAAAATTAATGCTCTCTCGATA 451
QY 103 eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123
Db 452 TAGAATAAATAAATGGTAAAGTAAATTTACCTGGATATATACTATAATTCATGTTAAAAA 511
QY 123 nasnSerGlyTrpSerIleGlyIleIleSerAsnPheIleValPheThrLeuLysGlnAs 143
Db 512 TAACTCAGGTTGGAGTATAGGTATATTAGTAATTTTATAGTATTTACTTTAAACAAAA 571
QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy 163
Db 572 TCAAGATAGTGAACAAAGTAAATTTTAGTTATGATATATCAAAATTAATGCTCTCGATA 631
QY 163 rAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl 183
Db 632 CAATAAATGGTTTTTTTGTAACGTGTACTTAACAATATGATGGGAAATATGAAGATTTATAT 691

```

```

QY 183 eAsnGlyLysLeuLeuAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerIly 203
DB 692 AAATGGAATAATATAGACTATATAAAGTAAAGAACTAACTGGAATAATTTTAGCAA 751
QY 203 sPhrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223
DB 752 AACTATAACATTGAAATAAATAAATCCAGATACCAGTTTGTATCTACATCTCTGA 811
QY 223 pAsnIleAsnMetTrpIleArgAspPheTrpIlePheAlaLysGluLeuAspGlyLysAs 243
DB 812 TAACATCAATATGTTGATAACAGATTTTATATATTTGCTAAAGAAATAGATGTAAGA 871
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnValValLysAspTrpIleGlyAs 263
DB 872 TATTAATATATTTTAACTGTCGCAATATTTTATATATTTGCTAAAGAAATAGATGTAAGA 931
QY 263 nAspLeuArgTrpAsnLysGluIleThrMetValAsnIleAspTrpIleAsnArgTrpIle 283
DB 932 TGATTTAAGATATATAAAGAAATATATGTTTAAATATAGATTTTAAATAGATATAT 991
QY 283 tTyrrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnG 303
DB 992 GTATCGCAACTCAGACAAATGTTTAAATACACGTAGAAATATAATGACTTCAATGA 1051
QY 303 uGlyTrpIleLysIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgG 323
DB 1052 AGCATATAAATATATAAAGAAATCAGAGAAATACAAATGATACAGATAGACGAGG 1111
QY 323 yGlyAspIleLeuTrpPheAspMetThrIleAsnAsnLysAlaTrpAsnLeuPheMetIly 343
DB 1112 AGGAGATATTTTATATTTGATATGACAAATATAAACAAGCATATAATTTGTTTGA 1171
QY 343 sAsnGluThrMetTrpAlaAspAsnHisSerThrGluAspIleTrpAlaIleGlyLeuAr 363
DB 1172 GAATGAACATGATGATGAGATATATCATGATCAGAGATATATATGCTATAGTTTAAG 1231
QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleCinProMetAsnAsnTh 383
DB 1232 AGAACAACAAGGATATAAATATGATATATATATATTTCAATACAAACCAATGAATAAC 1291
QY 383 rTyrrTrpTrpAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyI 403
DB 1292 TTATATTATACGATCTCAATATTTAAATCAATTTTAAATGAGAGAAATATTTCTGGAAT 1351
QY 403 eCysSerIleGlyThrTrpArgPheArgLeuGlyGlyAspTrpTrpArgHisAsnTrpLe 423
DB 1352 ATGTTCAATAGTACTTATCGTTTATAGACTTGGAGGTGATTTGATATAGACACAAATATT 1411
QY 423 uValProThrValLysGlnGlyAsnTrpAlaSerLeuLeuGluSerThrSerThrHisTr 443
DB 1412 GTGCTACTGTGAAGCAAGGAATATGCTTCATTTATAGAAATCAACATCAACTCATTTG 1471
QY 443 pGlyPheValProValSerGlu 450
DB 1472 GGGTTTGTACCTGTAAGTGAA 1493

RESULT 5
ID AA298630 standard; DNA; 1208 BP.
XX
AC AA298630;
XX
DT 20-JUN-2000 (first entry)
XX
DE C. botulinum type D toxin nucleotide sequence.
XX
KW Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
XX protect; ds.
XX Clostridium botulinum.
XX
PN WO200005252-A1.

```

```

XX 03-FEB-2000.
XX 20-JUL-1999; 99WO-IB01301.
XX 22-JUL-1998; 98ZA-0006538.
XX (AGRI-) AGRIC RES COUNCIL.
XX De Bruyn EE, Botha AD;
XX WPI: 2000-205375/18.
XX P-PSDB; AAY78982.
XX Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin
XX type D, useful in vaccines for protection against botulism, comprises
XX at least one amino acid mutation not present in the wild type D
XX neurotoxins
XX Claim 8; Page 54-57; 66pp; English.
XX This sequence represents a nucleotide sequence of a synthetic gene
XX encoding a non-toxic immunogenic derivative of Clostridium botulinum type
XX D toxin (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle
XX and sheep, and usually results in the death of the affected or poisoned
XX animal. The non-toxic immunogenic fragments of the C. botulinum
XX neurotoxin are useful in vaccines to protect animals (e.g. humans,
XX cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic
XX fragments can be produced relatively simply and inexpensively
XX (specifically by fermentation techniques). As the fragments are not
XX toxic the risk to production staff is reduced.
XX Sequence 1208 BP; 416 A; 203 C; 216 G; 373 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 1,43e-133 Length: 1208
Score: 1599.00 Matches: 307
Percent Similarity: 84.16% Conservative: 33
Best Local Similarity: 75.99% Mismatches: 56
Query Match: 66.46% Indels: 8
DB: Gaps: 4

```

US-09-910-186A-10 (1-450) x AA298630 (1-1208)

```

QY 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLys 69
DB 4 GCAGAAGTAGAGTTGAGGGAAACGTCAGTGAATCCTATCTTCATTGACTTTAG 63
QY 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89
DB 64 TTGGTAGCTCTGCTGATGACCGTGGCAAAATATTGTAACCAAAACGAAATATTGT 123
QY 90 TyrAsnSerMetTrpGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpVal 109
DB 124 TATAATGCTATGATGATCTTCTCTATTAGCTTCTGGATCAGAAATATAAATGGGT 183
QY 110 SerAsnLeuProGlyTrpTrpIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129
DB 184 TCAATTTACCAGGTTATACCATCATCGACAGCGTTAAATTAATTACGATGCTATT 243
QY 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149
DB 244 GGCATAATTAGCAATTTTCTAGTTTACACTAAAGCAAAACGAAATACGAGCAGGAC 303
QY 150 IleAsnPheSerTrpAspIleSerAsnAsnAlaProGlyTrpAsnLysTrpPhePheVal 169
DB 304 ATTAATTTTCTTAGCATTTTCAAGAACGCTGCTGCTGTTATATAAATGGTCTTTGA 363
QY 170 ThrValThrAsnAsnMetMetGlyAsnMetLysIleTrpIleAsnGlyLysLeuIleAsp 189
DB 364 ACCATTACCACAAACATGATGGTAAATATGATGATCATCTACATTAACGAAACGTAATCGAC 423
QY 190 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209

```

Db	424	ACCATCAAAAGTGAAGAATTGACCGTATCAACTTCTCAAGACCATTACGTTTCAGATG	483
QY	210	AsnLysIleProAsnThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle	229
Db	484	ATAAAATCCCAACACAGGATTAATAACGCTGCTGATTCTGACACACATCAATATGTGGATT	543
QY	230	ArgAspPheTrpIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
Db	544	ACAGATTTCCTATATTTTCGAAAAGAAATACATGACAGGACGACATATATATCCCTTCAT	603
QY	250	SerLeuGlnTrpThrAsnValIleValIleAspTrpTrpGlyAsnAspLeuArgTyrAsnLys	269
Db	604	TCATTTCATACACAAATGTTGTTAAGGATTATTGGGGTAATGATGATTAAGATATGATAAG	663
QY	270	GluTyrTrpMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGln	289
Db	664	GAGTACTACATGATTAATGTTAAATTACATGAATCGTTACATGCTCTAAAAGGGTACGGT	723
QY	290	IleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIle	309
Db	724	ATCGTCTTCAACACACGTAATAAATAACACGATTTCACGAAAGGGTACAAAATATCATATA	783
QY	310	LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe	329
Db	784	AAAAGAATCCGTGGAAATACAAATGATACTAGATTCGTGGTGAGACGTCGTATATTTT	843
QY	330	AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla	349
Db	844	ATAACACCATAGATAAATACGACGACTCTTTA-----GGCATGTATAAA	888
QY	350	AspAsnHisSerThr---GluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAsp	368
Db	889	CCGTCAAGAAACTTAGGACACAGACCTGCTCCATTTGGGTGCATTAGATCAACCGATGGAC	948
QY	369	---IleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAla	387
Db	949	GAAATTCGTAATATGGATCATTCATTATCAACCATGTAATACGTTTCGATTACTACGGC	1008
QY	388	SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly	407
Db	1009	AGCCAATTATTTCTTTCTAGCAACCCAGCACCAATCGTTTAGGAATACATCAATCGGC	1068
QY	408	ThrTyrArgPheArgLeuGlyClyAsp--TrpTyrArgHisAsnTyrLeuValProThr	426
Db	1069	TCATCTCTTTAACTGGGTGATGATTACTGGTTTAAACCAATGAATCTTGATCCCCAGTA	1128
QY	427	ValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheVal	446
Db	1129	ATTAATAATCGAACATACGCTTCATTACTTGGAAATCAACTTCTACACATTTGGTCTTGTG	1188
QY	447	ProValSerGlu	450
Db	1189	CCAGCGCTCGAG	1200
RESULT	6		
AAZ98631			
ID	AAZ98631	standard; DNA; 1440 Bp.	
XX	AAZ98631;		
AC			
AC			
DT	20-JUN-2000	(first entry)	
XX		Type D BoNT non-toxic immunogenic derivative nucleotide sequence.	
DE		Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison	
KW		protect; ds.	
XX			
OS		Clostridium botulinum.	
XX			
PN		W0200005252-A1.	
XX			
PD		03-FEB-2000.	

20-JUL-1999; 99WO-IB01301.  
22-JUL-1998; 98ZA-0006538.  
(AGRI-) AGRIC RES COUNCIL.  
De Bruyn EE, Botha AD;  
WPI; 2000-205375/18.  
P-PSDB; AAV78982.  
Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin type D, useful in vaccines for protection against botulism, comprises at least one amino acid mutation not present in the wild type D neurotoxins -  
Claim 8; Page 59-62; 66pp; English.  
This sequence represents a nucleotide sequence of a synthetic gene encoding a non-toxic immunogenic derivative of Clostridium botulinum type D toxin (BoNT). The sequence includes portions of a suitable plasmid which are immediately upstream and downstream of the gene. Botulinum neurotoxin causes botulism poisoning in cattle and sheep, and usually results in the death of the affected or poisoned animal. The non-toxic immunogenic fragments of the C. botulinum neurotoxin are useful in vaccines to protect animals (e.g. humans, cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic fragments can be produced relatively simply and inexpensively (specifically by fermentation techniques). As the fragments are not toxic the risk to production staff is reduced.

XX  
SQ  
Sequence 1440 BP; 477 A; 262 C; 272 G; 429 T; 0 other;

Alignment Scores:		
Pred. No.:	1.77e-133	Length:
Score:	1589.00	Matches:
Percent Similarity:	84.16%	Conservative:
Best Local Similarity:	75.99%	Mismatches:
Query Match:	66.46%	Indels:
DB:	21	Gaps:
		4
		8
		1440
		307

US-09-910-186A-10 (1-450) x AAZ98631 (1-1440)

Ov 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePropheAspPheLys 69

Db 58 GCAGAAAGTTAGAGTTGAGGGAAACGTCAGTTGAATCCCTATCTTCCATTGACTTTAAG 117

Qy 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89

Db 118 TTGGGTAGCTCTGGTGATGACCGTGGCAAAATATTGTAACTCAAAACGAAATATTGTT 177

Qy 90 TyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpVal 109

Db 178 TATAATGCTATGATGAGTCTTTCTCTCTATTAGCTTCTGGATCAGAAATTAATAATGGGTT 237

Qy SerAsnLeuProGlyTyrThrIleLeuAspSerValLysAsnAsnSerGlyTrpSerIle 12

DB 238 TCAGATTTACCAAGGTATATACCATCATCGACAGCGTTAAAAAATTAATTCAGGATGGCTGTA 239

[illegible][illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1980	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																					

Qv 190 ThrIleIysValIysGluLeuThrGluValLeuAsnPheserIysThrIleThrPheGluIle 209

[illegible]

```

Db 478 ACCATCAAGTGAAGATTCACCGGTATCAACTTCTCAAGACCATTTACGTTTCAGATG 537
Qy 210 AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetIrpIle 229
Dd 538 AATAAAATCCGACACAGCATTAATAACGCTCTGATCTGCACAAATCAATATATGCGATT 597
Qy 230 ArgAspPheIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn 249
Dd 598 AGAGATTTCTATATTTCCGAAAGAAATAGATGACAGACATTAATATCTCTTCAT 657
Qy 250 SerLeuGlnIleThrAsnValLysAspIleTyrIlePheValAspLeuArgIleAsnLys 269
Dd 658 TCACCTCAGTACATAATATGTTGTAAGGATTTATGGGTAATGATTAAGATATGATAAG 717
Qy 270 GluIleIleThrMetValAsnIleAspIleLeuAspArgIleMetIleAlaAsnSerArgGln 289
Dd 718 GAGTACTACATGATTAATGTAATATCATGATTCGTTACATGCTGCTAAAAGGGTAAACGGT 777
Qy 290 IleValPheAsnThrArgIleAsnAsnAspPheAsnGluGlyIleIleIleIle 309
Dd 778 ATCGTCTTCAACACTCGTAAATAATAACAACGATTTCAACGAAGGGTACAAAATATCATATA 837
Qy 310 LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuIlePhe 329
Dd 838 AAAAGATCCGTGGAAATACAAATGATATGATGAGTTCGGTGGAGAACGTCGTATATTT 897
Qy 330 AspMetThrIleAsnAsnLysAlaIleTyrAsnLeuPheMetLysAsnGluIleMetIleIle 349
Dd 898 AATCAACGATAGATAATAAGCAGTACTCTTA-----GGCATGTATATAA 942
Qy 350 AspAsnHisSerThr---GluAspIleTyrAlaIleGlyLeuArgGluGlnIleLysAsp 368
Dd 943 CCGTCAAGAACTPAGGACAGACCTTTGTCATTTGGGTGCATTAAGATCAACCGATGGAC 1002
Qy 369 ---IleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrIleIleIleIle 387
Dd 1003 GAATTCGTAATATGATCATTTATTTCAACCATGTAATACGTTCCGATTCATACGGC 1062
Qy 388 SerGlnIlePheLysSerAsnPheAsnGlyLysIleSerGlyIleCysSerIleGly 407
Dd 1063 AGCCAAATATTTCTTCTAGCAAGCGCACCAACATCGTTTGAAGATATCATCAATCGCGC 1122
Qy 408 ThrIleArgPheArgLeuGlyGlyAsp---TyrIleArgHisAsnIleValProThr 426
Dd 1123 TCATATCTTTTAACATGGGTGATGATCTGCTTAAACCATGAATACATGATCCAGTA 1182
Qy 427 ValLysGlnGlyAsnIleTyrAlaSerLeuLeuGluSerThrSerThrHisIlePheVal 446
Dd 1183 ATTAATAATCGAACATTAACGCTTTCATTAAGTAAATCAACTTCTACATTTGGTCTTGTG 1242
Qy 447 ProValSerGlu 450
Dd 1243 CCAGGCTGAG 1254
RESULT 7
AAA54487
ID AAA54487 standard; DNA; 1374 BP.
XX
AC AAA54487;
XX
DT 11-APR-2001 (first entry)
DE
DE Botulinum toxin heavy chain C-terminal coding sequence (serotype D).
KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
FH key Location/Qualifiers
FT CDS 10..1365

```

```

FT
FT /*tag= a
   /*product= H_C peptide fragment

```

WC200067700-A2.

16-NOV-2000.

12-MAY-2000; 2000WO-US12890.

12-MAY-1999; 99US-0133865.

12-MAY-1999; 99US-0133866.

12-MAY-1999; 99US-0133867.

12-MAY-1999; 99US-0133868.

12-MAY-1999; 99US-0133869.

12-MAY-1999; 99US-0133873.

29-JUL-1999; 99US-0146192.

(USSA ) US ARMY MEDICAL RES &amp; MATERIAL COMMAND.

Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

P-PSDB; AAB04093.

WPI: 2001-016048/02.

P-PSDB; AAB04093.

New nucleic acids encoding the carboxy- or amino-terminal portions of

the heavy chain of botulinum neurotoxin of serotype A-G, useful as

vaccine against botulism

Claim 2; Fig 6a; 73pp; English.

Botulinum neurotoxins are translated as a single 150 kDa polypeptide

chain and then posttranslationally nicked, forming a dichain

consisting of a 100 kDa heavy chain and a 50 kDa light chain which

remain linked by a disulfide bond. Nucleic acids encoding the

carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy

chain of botulinum neurotoxin (BoNT) can be used in recombinant

expression vectors and expressed in transformed cells to produce

peptide antigens useful for eliciting an immune response to give

protective immunity against botulinum neurotoxin, which causes

botulism. The nucleic acids are expressible in a recombinant

organisms such as Escherichia coli or Pichia pastoris. The use

of recombinant nucleic acids are advantageous since it eliminates

the need to culture large quantities of hazardous toxin-producing

bacterium, production yield from the genetically engineered product

is also high and cost of production is lower. The nucleic acids can

be derived from Clostridium botulinum serotypes A-G.

XX Sequence 1374 BP; 367 A; 423 C; 274 G; 310 T; 0 other;

Alignment Scores:

Pred. No.: 9,38e-69 Length: 1374

Score: 868.50 Matches: 197

Percent Similarity: 62.06% Conservative: 86

Best Local Similarity: 43.20% Mismatches: 145

Query Match: 36.32% Indels: 28

DB: 22 Gaps: 11

US-09-910-186A-10 (1-450) x AAA54487 (1-1374)

Qy 2 ThrIleProPheAsnIlePheSerIleThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21

Dd 49 ACCAAGCATTCACATCTCTCTCTACACCAACAACTCCTGTTGAGGACATCATCAAC 108

Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41

Dd 109 GAGTACTCAACTCCATCAACGACTCCAGATCTTCTCCTTCAGAACCAAGAAAGACGCC 168

Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61

Dd 169 TTGTCGACACCTCCGTTTAAACCGAGGTGAGTCCGTCACACACGTCCTCAGTGAAC 228

Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81

FT



Db 229 ACCATCTACACCAAGACTTCAAGTTGCTCTCCGGTGAC-----AGATCATC 279  
QY 82 ValThrGlnAsnGluAsnValThrAsnSerMetTyrGluSerPheSerPhe 101  
Db 280 GTCAACTTGAAACAACAACATCTGTACTCCGCCATCTACGAAACTCCTGTCTCCTC 339  
QY 102 TPIleArgGlnAsnLysTrpValSerAsnLeuProGly---TyrThrIleLeuAspSer 120  
Db 340 TGGATCAGATCTCCAGAGCTTGACCACTTGCCACAAAGAGTACACCACTCACTCACTC 399  
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleLeuValPheLeuThrLeu 140  
Db 400 ATCGAGCAGAACTCCGGTTGGAAGTTGTTATCCGTAACGGTAACATCGAGTGGATCTTG 459  
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160  
Db 460 CAGGACGTCACCGTAAGTACAGCTGCTGATCTTCGACTCCGAGTCTCTGCCAC 519  
QY 161 ProGlyTyr---AsnLysTrpPheValThrValThrAsnMetMetGlyAsnMet 179  
Db 520 ACCGGTTACACCAACAAGTGTCTCTCGTACCATCCACCAACATCATGGTTACATG 579  
QY 180 LysIleTyrIleAsnGlyLysLeuLeuAspThrIleLysValLysGluLeuThrGlyLe 199  
Db 580 AAGTTGTATCATCAGCGGTGAGTTGAAGCAGTCCAGAGATCGAGGACCTGGACGAGTC 639  
QY 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219  
Db 640 AAGCTGGACAAGACCATCTCTCGGTATCCAGGACATCGAC----- 684  
QY 220 SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239  
Db 685 -----GAGAACCAAGATGTTGGATCGTGACTCAACATCTCTCCAAAGGAGCTG 735  
QY 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAsp 259  
Db 736 TCCACGAGGACATCAATGCTGTACAGGGTCAGATCTCGAGAACGTCATCAAGGAC 795  
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279  
Db 796 TACTGGGTAAACCACTGAAGTTCACACCGAGTACTACATCATCAACGACAACTACATC 855  
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsn 299  
Db 856 GACCGTTATCTCCGCCAGGAGCAACGTCGTGTGTCTGTGTCTGTGTCTGTGTCTGT 915  
QY 300 AspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThr 319  
Db 916 AAGCTGTACACCGGTAAACCTATCACCATCAAGTCCGTCCTCGACAAAGACCTTACTCC 975  
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339  
Db 976 CGTATCTCTGAACGGTGACAACTATCATCTGTCATCTGTGTGTGTGTGTGTGTGTGT 1035  
QY 340 LeuPheMetLysAsnGluThrMetTyrAla-----AspAsnHisSerThrGluAsp 356  
Db 1036 ATCATCCGTGACACCGACACCATCTACCCACCGAGGTGTGTGTGTGTGTGTGTGTGT 1095  
QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376  
Db 1096 GTCTACGCCCTGAAGCTGACGTCCAACTGCGGTGATCTACGATCGGTATCTCTCC 1152  
QY 377 IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn 396  
Db 1153 ATCAAGACATCTGTCTCAAGNACAGTGTCTCCAGATCTTC---TCTCTCTCCGT 1209  
QY 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly 415  
Db 1210 ---GAGAACACCATGCTGCTGCGACATCTACAGCCTTGGCGTTTCTCTCC----- 1257  
QY 416 AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435  
Db 1258 -----TTCAAGAACCGCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1305

QY 436 LeuGluSerThrSerThrHisTrpGlyPheVal-ProValSerGlu 450  
Db 1306 CTGCTGTCCACTCCTCTCTTCTGGAAGTTATCTCTCCGTTGCCAG 1351

## RESULT 8

AAV30591 standard; DNA; 1469 BP.

XX AAV30591;

XX 07-DEC-1998 (first entry)

XX Clostridium botulinum type D toxin C fragment gene in pETHisb.

XX Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;  
KW botulism; BotD; ds.

XX Clostridium botulinum serotype D.  
OS Synthetic.

XX Key Location/Qualifiers  
CDS 108..1463  
/\*tag= a

XX WO9808540-A1.

XX 05-MAR-1998.

XX 28-AUG-1997; 97WO-US15394.

XX 28-AUG-1996; 96US-0704159.

XX (OPHI-) OPHIDIAN PHARM INC.

XX Thalley BS, Williams JA;

XX WPI; 1998-230234/20.

XX P-PSDB; AAW68398.

XX Host cell containing recombinant expression vector encoding  
PT Clostridium botulinum type B or E toxin - useful to treat humans  
PT and other animals at risk of intoxication with clostridial toxin

XX Example 47; Page 352-53; 428pp; English.

XX This is the DNA sequence of the Clostridium botulinum serotype D  
CC neurotoxin fragment C gene contained in plasmid pETHisb. The  
CC encoded BotD fragment C polypeptide (see AAW68398) has a His-tagged  
CC N-terminal extension. The vector can be used to express native  
CC soluble BotD C fragment in Escherichia coli host cells. The  
CC invention relates to recombinant proteins derived from C. botulinum  
CC toxins. Methods are provided which allow for the isolation of  
CC soluble recombinant clostridial toxins free of significant  
CC endotoxin contamination. Preferred hosts are E. coli, insect cells  
CC and yeast cells. The recombinant toxins are used as immunogens for  
CC the production of vaccines and antitoxins that are useful in the  
CC treatment of humans and animals at risk of intoxication with  
CC clostridial toxin.

XX Sequence 1469 BP; 581 A; 146 C; 245 G; 497 T; 0 other;

## Alignment Scores:

Pred. No.: 7,06e-63 Length: 1469  
Score: 803.50 Matches: 186  
Percent Similarity: 60.22% Conservative: 85  
Best Local Similarity: 41.33% Mismatches: 151  
Query Match: 33.61% Indels: 28  
DB: 19 Gaps: 11

US-09-910-186A-10 (1-450) x AAV30591 (1-1469)

QY 4 PropheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleAsnGluTyr 23



```

Db 152 CCATATCGAAGGTCCTCATATGCGTAGCATGCTTTATTAAGAGATATATTTATGATGATA 211
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 212 TTTCAATAGTATTAATGATTCACAAATTTTGAGCTTACAAACAAAAAATGCTTAGT 271
QY 43 lAspThrSerGlyTyrrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProil 63
Db 272 GCATACATCAGGATATATGAGAGAGTGGGTAGGAGATATGTTCACTTATACGAT 331
QY 63 ePheProPheAspPheLysLeuGlySerSerGlyLysAspArgGlyLysValIleValTh 83
Db 332 ATATACAAATGACTTTAAATTAAGTAGTTCAGGAGAT-----AAAATTAAGTAA 382
QY 83 rGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPheThrpi 103
Db 383 TTTAAATATATATTTATATAGCGCTATTTATGAGAACTCTAGTGTAGTTTGGAT 442
QY 103 eArgIleAsnLysTrpValSerAsnLeuProGly---TyrrThrIleIleAspSerVally 122
Db 443 TAAGATATCTAAAGATTTAACTAATTCCTCATATGAATGATATACAAATTAACAGTATAGA 502
QY 122 sAsnAsnSerGlyTyrrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysG1 142
Db 503 ACAAAATCTGGTGGTGAATATGATTAGTAGTGGCAATATAGAAATGGATTTTACAGA 562
QY 142 nAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAsnAlaProG1 162
Db 563 TGTTAATAGAAAGTATAAAGTTTAAATTTTATGATTATAGTGAATCATTAAGTCATACAGG 622
QY 162 YTyrr---AsnLysTyrrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys11 181
Db 623 ATATACAAATTAATGGTTTGTGTACTATACCAATTAATATATATATGAGGTATATGAAACT 682
QY 181 eTyrrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPh 201
Db 683 TTATATAAATGGAGAATTAAGCAGCAGCAAGAAATTCAGATTTAGATGAGGTAAAGTT 742
QY 201 eSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAs 221
Db 743 AGATAAAACCATAGTATTTGGAAATAGATGAGAAATATAGAT----- 782
QY 221 pSerAspAsnIleAsnMetTyrrIleArgAspPheTyrrIlePheAlaLysGluLeuAspG1 241
Db 783 ----GAGAAATCAGATCCTTGGATTAGAGATTTAAATATTTTCTTAAGAAATTAAGTAA 838
QY 241 yLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrrThrAsnValValLysAspTyrr 261
Db 839 TGAAGATATTAATATGTTATATAGGGGCAAAATATTAAGAAATGTTTAAAGATTATG 838
QY 261 pGlyAsnAspLeuArgTyrrAsnLysGluTyrrMetValAsnIleAspTyrrLeuAsnAr 281
Db 899 GGGAAATCCTTTGAAGTTGTACAGATATATATATATATATATATATATATATATAGATAG 958
QY 281 gTyrrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPh 301
Db 959 GTATATTGCACCTGAAAGTAACTTACTTGTCTGTTGCGGTATCCAGATAGATCTAAAT 1018
QY 301 eAsnGluGlyTyrrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVa 321
Db 1019 ATACTGGAATCCTATATCTATTTAATCTAGTATCTCATTAAGAAATCCTTATAGAAAT 1078
QY 321 lArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPh 341
Db 1079 TTTAAATGGAGATAATAATCTTCATATGTTATATATAGTAGAGAAATATATAGTAAT 1138
QY 341 eMetLysAsnGluThrMetTyrrAlaAspAsnHis-----SerThrGluAspIleTy 358
Db 1139 AAGAGACTACTGATACAAATATATGCAACACAGAGGAGAGGTCTTCACAAAATTTGTGATA 1198
QY 358 rAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleG1 378

```

```

Db 1199 TGCATTAAATACAGAGTAATTTAGGTATTTAGTATAGT---ATATTAGTATATA 1255
QY 378 nProMetAsnAsnThrTyrrTyrrAlaSerGlnIlePheLysSerAsnPheAsnGlyG1 398
Db 1256 AAATATTGTTATCTTAAATAAATAATATTTAGTCAAAATTTTC---TCTAGTTTATAGG---GA 1309
QY 398 uAsnIleSerGlyIleCysSerIle---GlyThrTyrrArgPheArgLeuGlyGlyAspTr 417
Db 1310 AAATACAAATGCTTCTAGCAGATATATATAACCTTGGAGATTTCT----- 1355
QY 417 pTyrrArgHisAsnTyrrLeuValProThrValLysGlnGlyAsnTyrrAlaSerLeuLeuG1 437
Db 1356 -TTTAAATAATCATACAGCCAGCTGTCAGTAACT-----AATATGAAACAAACTATT 1408
QY 437 uSerThrSerThrIstTrpGlyPheVal 446
Db 1409 ATCACTTCATCTTTTGGAAATTTATT 1436
RESULT 9
AAZ87215
ID AAZ87215 standard; DNA; 1400 BP.
XX AAZ87215;
XX 08-MAY-2000 (first entry)
XX DNA encoding synthetic BoNT serotype E (BoNTE) Hc fragment.
DE BoNT neurotoxin; heavy chain; BoNT; serotype E;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
XX VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX Clostridium botulinum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 10..1359
FT /tag= a
FT /product= "Synthetic botulinum neurotoxin serotype E
(X) (BoNTE) heavy chain C-terminal fragment (Hc)"
XX
PN WO200002524-A2.
XX
XX 20-JAN-2000.
XX 09-JUL-1999; 99WO-US15570.
XX
XX 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
WPI: 2000-160827/14.
DR P-PSDB; AAY77137.
XX
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
toxin serotypes A-G, is used for inducing an immune response against
botulinum.
XX
XX Disclosure; Page 43-44; 54pp; English.
XX
XX The invention relates to novel vaccines that induce a protective immune
response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
DNA construct comprising a vector, and at least one nucleic acid
fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
serotypes A-G. In preferred embodiments of the invention, the vector is
a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
this vector results in the production of large amounts of a protein
encoded by a sequence cloned into the replicon. The constructs are used
to produce vaccines against botulism. The proteins can also be used as

```

diagnostic tools for the diagnosis of botulism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine of overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AA82712-787217 represent synthetic DNA sequences encoding PONT HC fragments used in the present invention. These were optimised for codon usage for expression in yeast.

SQ Sequence 1400 BP; 483 A; 298 C; 232 G; 387 T; 0 other;

Alignment Scores:

Pred. No.:	1.13e-43	Length:	1400
Score:	589.50	Matches:	152
Percent Similarity:	51.93%	Conservative:	90
Best Local Similarity:	32.62%	Mismatches:	159
Query Match:	24.65%	Indels:	65
DB:	21	Gaps:	17

US-09-910-186A-10 (1-450) x AAZ87215 (1-1400)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21

Db  
64 AGTATTCCCTTTTAAGCTTTCTTCTTATACAGATGATAAAATTTAATTTCTACTTCAAC 123

QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41

Db 124 AAGTCTTCAAGAGAAATTAAGTCTTCTCCGTTTAAACATGAGATACAAGAAATGATAA 183

42 LeuValAspThrSerGlyTyrAsnAlaGlyValSerGlyGlyGlyValGlyIleAsp 61

D<sub>b</sub> 184 TACGTCGACACTTCGGGTTACGACTCCAAATATCAACATTAA CGGTGACGTACAAGTAC 243

62 proIlephepropheAsppheTysIeugIyserSerGIVGIIAspAraGIVIysValIle 81

D<sub>b</sub> 244 CCAACTAACAAAACCAATTC-----GGTATCTACAACGACAAGCTTACTGAGCTGAAC 297

82 ValThrGlnAsnGlnAsnTleValTyrAsnSerMetTyrGlnSerPheTleSerPheTle

298 ATCTCTCAAAACGACCTACACTTATCTACGACACACGACCTCTCTTCTTTC 357

102 TrnIleArgIle-----AsnValSerAsnLeuProClvTyrThrIle 117

[illegible]

Oz: 118 TleXenSorValIwC=...XnXnSOrTlOc]vTlOtJcSOrXnRPhOv 136

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

<

0v 127 V<sub>a</sub>] pbeThrIeI'ncQ] nAenC] uDnScRQ] uG] nSgRT] eAnpbScRE.u.nZcNT] c 156

[illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible]

7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530

4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 5

[illegible]

These new units of production have the potential to be a

Db	751	GACAGGAGTTAGACGAGACCGGAGATTCAAACCTTTATACAGCAACGCAACCTTAACCAAT	810
Qy	256	valvallylsaspytirpGlyAsnaspLeuAargTyrAsnLysGluTyrTyrMetValasn	275
Db	811	ATTTTGAAGGACATCTGGGTAACTACTTTCCTTTCAGCAAGGAATACTACTATTAAAC	870
Qy	276	ileAspTyrLeuasnAargTyrMet-----TyrAlaasnSerAargGlnileValPheasn	293
Db	871	GTGTAAAGCCAAACAACACTTCATTGATAGGAAAGGATTCTTACTTTAAGCATTAAACAAC	930
Qy	294	ThrAargArgasn-----AsnAsnaspPheasnGluGlyTyrLysileileileLys	310
Db	931	ATCAGAAAGCACTATCTTTAGCTACAGATATTAATCTCTGGTATCAAGGTAAAGATCCAA	990
Qy	311	ArgIle---ArgGlyAsnThrAsnAspThrAargValAargGlyGlyAspLileLeuTyrPhe	329
Db	991	AGAGTTAAACAACCTCTCTACTAACGATAACCTTTGTAGAAAGACGATCAGTCTATATT	1050
Qy	330	AspMetThrIleasnAsnLysAlaTyrAsnLeuPheMetLysasnGluThrMetTyrAla	349
Db	1051	AACTTCGTCGCTAGCAAGACT-----CACTTATTC-----CCATTATATGCT	1092
Qy	350	AspAsnHisSerThrGluAspIleTyrAlaileGlyLeuAargGluGlnThrLysAspile	369
Db	1093	GATACCGCTACCACCAAC-----AAGGAGAAGACCATCAAGATC	1131
Qy	370	AsnAsp-----AsnIlellePheGlnIleGlnPrometAsnAsnThrTyrTyrTyrAla	387
Db	1132	TCCCTCCTCGCAGACAGATTAAACCAAGTCGCTGTATGAACCTCGTCGGTAAACAAGTGT	1191
Qy	388	SerGlnIlePheLysSerAsnAsnAsnGlyGluAsnIleSerGlyIleCysSerIleely	407
Db	1192	ACCATGAACCTTTAAAAATAAT---AATGGAATAAT-----ATTGGG	1230
Qy	408	ThrTyrAargPhearg-----LeuGlyGlyAspTrpTyrAargHisasnTyrLeu	423
Db	1231	TTGTTAGTTTCAAGGCAGATAGTGTAGTGTAGTACTTGTAT-----	1275
Qy	424	ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHis---	442
Db	1276	-----TATACCCACATGAGAGATCACACCAAGCAAGCAATGGA	1311
Qy	443	-----TrpGlyPheVal 446	
Db	1312	TGTTTTTGGAACTTTATT 1329	
RESULT	10		
AAA54488			
ID	AAA54488		
XX	AAA54488	standard; DNA; 1400 BP.	
AC	AAA54488;		
XX			
DT	11-APR-2001	(first entry)	
XX			
DE			
XX			
KW	Botulism toxin heavy chain C-terminal coding sequence (serotype E).		
KW	Botulism; toxin; neurotoxin; heavy chain; recombinant expression;		
KW	recombinant vector; antigen; immune response; vaccine; bacterium;		
KW	infection; ds.		
XX			
OS	Synthetic.		
OS	Clostridium botulinum.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	10..1358	
FT		/*tag= a	
FT		/product= H_C peptide fragment	
XX			
EN	W0200067700-A2.		
XX			
PD	16-NOV-2000.		

8757







Db 3286 ATTTTAAAGACTTTTGGGTGATATTACAAATATGATAAACCACATATATGTTAAAT 3345  
 Qy 276 IleAspTyrIleuAsnArgTyrMetTyrAlaAsnSer- 287  
 Db 3346 TTATATGATCAATAAATATGTCGATGTAATATGTTAGGTATATGATGAT 3405  
 Qy 298 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302  
 Db 3406 CTTAAAGGGCTAGAGGTAGGTAATGACTACAAACATTTATTTAAATCAAGTTGTAT 3465  
 Qy 303 GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322  
 Db 3456 AGGGGCAAAATTTATTAATAAAATATGCTCTGGAAATTAAGATATATTTGTTAGA 3525  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3526 AATAATGATCGTATATATTAATGATAGTTAAATAAAGATATAGTTA----- 3579  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3580 -----GCTACTAATGCTACACAGCGGCGTAGAAAATACTAAGT 3621  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3622 GCATTAGAAATACCTGATAGGAATCTAAGTCAAGTAGTATAGTCAAGAAAT 3681  
 Qy 383 ThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 3682 GATCAAGGAATPACAAATAATATGCT---AAATGAATTAACAAGATATAATGGGATGAT 3738  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsnTyr 422  
 Db 3739 ATAGGCTTTATAGGATTCATCAGTTI-----AATAAT 3771  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
 Db 3772 ATAGCTAACTAGTAGCAAGTAATTTGTTATTAATAGACAAATAGAAAGATCTAGTAGGACT 3831  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 3832 TTGGGTTCCTCATGGGAATTTATTCCTGTAGATGAT 3867  
 RESULT 13  
 AAC64582  
 ID AAC64582 standard; cdna; 4835 BP.  
 XX  
 AC AAC64582;  
 XX  
 DT 15-FEB-2001 (first entry)  
 DE BoNT/A neurotoxin prototoxin nucleotide sequence SEQ ID NO:10.  
 XX  
 KW Human; procholecystokinin; CCK A receptor; CCK B receptor;  
 KW pancreatitis; antiinflammatory; ss.  
 XX  
 OS Clostridium botulinum.  
 XX  
 PN WO200061192-A2.  
 XX  
 PD 19-OCT-2000.  
 XX  
 PF 06-APR-2000; 2000WO-US09142.  
 XX  
 PR 08-APR-1999; 99US-0288326.  
 XX  
 PA (ALLR ) ALLERGAN SALES INC.  
 PI Steward LE, Sachs G, Aoki KR;  
 XX  
 XX WPI; 2000-679416/66.  
 DR  
 PT New composition for treating acute pancreatitis, comprises a pancreatic  
 PT cell surface marker binding element, a translocation element that

PT transfers polypeptide across vesicular membrane, and a therapeutic  
 PT element  
 XX  
 PS Disclosure; Page 29-32; 50pp; English.  
 XX  
 CC The present invention describes a composition (I) for treating acute  
 CC pancreatitis. (I) comprises a first element containing a binding element  
 CC that binds to a pancreatic cell surface marker, a second element  
 CC containing a translocation element that facilitates polypeptide transfer  
 CC across the vesicular membrane, and a third element containing a  
 CC therapeutic element that inhibits enzyme secretion in pancreatic cell  
 CC cytoplasm. Also described is a method for making a therapeutic  
 CC polypeptide having a binding element selective for cholecystokinin (CCK)  
 CC receptor by expressing within a host cell a recombinant chimeric  
 CC polypeptide comprising an extein containing a therapeutic element and a  
 CC translocational element, and an intein located to the carboxy terminal  
 CC of extein having a cysteine, serine or threonine at its amino terminus,  
 CC and contacting the extein with a synthetic peptide comprising a CCK  
 CC amino acid sequence containing an amidated phenylalanine at a natural  
 CC C-terminus, and a cysteine, serine or threonine at its N-terminus, and  
 CC a nucleophilic reagent able to cause cleavage of the intein to form a  
 CC peptide bond between the extein C-terminus and synthetic peptide  
 CC N-terminus through the formation of an activated ester or thio ester  
 CC of pancreatic digestive enzymes, and prevents exocytic fusion of vesicles  
 CC containing secretory enzymes of pancreas. (I) is useful for treating  
 CC acute pancreatitis. The present sequence encodes the Clostridium  
 CC botulinum BoNT/A neurotoxin prototoxin which is given in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 4835 BP; 1934 A; 517 C; 756 G; 1628 T; 0 other;

Alignment Scores:  
 Pred. No.: 9,45e-43 Length: 4835  
 Score: 586.50 Matches: 149  
 Percent Similarity: 48.94% Conservative: 82  
 Best Local Similarity: 31.57% Mismatches: 186  
 Query Match: 24.53% Indels: 55  
 DB: 21 Gaps: 11

US-09-910-186A-10 (1-450) x AAC64582 (1-4835)

Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu 22  
 Db 2902 ATACCTTTTCAGCTTTCCAAATACGTAGATAATCAAGATATTATCTACATTACTGAA 2961  
 Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThrLeu 42  
 Db 2962 TATATTAGAAATATTATTAATCTCTATTTGAATTTAGATATGAAAGTAACTATTA 3021  
 Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
 Db 3022 ATAGACTTATCTAGGTATGCATCAAAATAAATATTGGTAGTAAAGTAAATTTTGATCCA 3081  
 Qy 53 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79  
 Db 3082 ATAGATAAAATCAAAATCAATTTATTTAATTTAGAAGTAGT-----AAA 3126  
 Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99  
 Db 3127 ATTGAGGTAAATTTAAAAATGCTATTGTATATATAGTATGTATGATAAATTTTAGTACT 3186  
 Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
 Db 3187 AGCTTTTGGATAAGAATTCCTAAGTATTTTAAACAGTAAAGTCTAAATAATGAATATACA 3246  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 3247 ATAATAATTCATGGAATAATATTCAGGATGGAAGTATCACTTAATATGCTGAATA 3306  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 3307 ATCTGGACTTACAGGATCTACAGGAATAAATAAAGAGTAGTATTTTAAATACAGCTCAA 3366

Qy 157 SerAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 Db 3367 ATGATTAATATATATATATAAAGAGATGATTTTGTAACTACTACTATAATAGA 3426  
 Qy 176 MetGlyAsnMetLysIleThrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 3427 TTAATAACTCTAAATTTATATAATGGAAGATTATAGATCAAAACCAATTCAAAT 3486  
 Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3487 TTAGGTAATATTCATGCTAGCTAATAATATAATGTTTAAATAGATGGTGTAGAGATACA 3546  
 Qy 216 GlyLeuIleThrSerAspSerAsnIleAsnMetTrpIleArgAspPheThrIlePhe 235  
 Db 3547 -----CATAGATATATTGGATAAAATATTAAATCTTTT 3582  
 Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTrpThrAsn 255  
 Db 3593 GATAAGGAATTAATGAAAGAAATCAAGATTTATATGATAATCAATCAATTCAGGT 3642  
 Qy 256 ValLysIleAspTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetValAsn 275  
 Db 3643 ATTTAAAGACTTTTCGGGTGATTATTACAATATGATAAACCATACTATATGTTAAAT 3702  
 Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
 Db 3703 TTATATGATCAATATAATATCGATGTAATAATAGTAGGTATATATGAT 3762  
 Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302  
 Db 3763 CTTAAAGGCGCTAGAGTAGCGTAATGACTACAAACATTTATTAAATCAAGTTTGTAT 3822  
 Qy 303 GluGlyTyrLysIleIleLysArgIleGlyAsnThrAsnAspThrArgValArg 322  
 Db 3823 AGGGGACAAATTTATATAAAATATGCTTCGGAATAAGATATAATTTGTAGA 3882  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3893 AATAATGATCGTGTATATATATGATAGTAGTAAATATAAGATATAGTTA----- 3936  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3937 -----GCTACTAATGCATCACAGCGCGGTAGAAAATACTAAGT 3978  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3979 GCATTAGAAATACCTGATGATAGGAAATCTAAGTCAAGTAGTAGTAATGAAGTCAAAAAAT 4038  
 Qy 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 4039 GATCAGGAATAACAATAATGTC---AAATCAATTTACAGATATAATGGGATGAT 4095  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyr 422  
 Db 4096 ATAGCGCTTATAGGATTCATCAGTTT-----AATAAT 4128  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
 Db 4129 ATAGCTAACTAGTACCAAGTAATTTGGTATTAATAGACAATAAGAGATCTAGTAGGACT 4188  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 4189 TTGGGTGCTCATGGGAATTTATTCCTGTAGATGAT 4224  
 RESULT 14  
 AAF58862  
 ID AAF58862 standard; DNA; 4017 BP.  
 AC  
 XX AAF58862;  
 XX  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX

DE C botulinum BONT/E neurotoxin cDNA.  
 XX Activatable neurotoxin; protease cleavage; H chain; L chain;  
 KW clostridial neurotoxin; ss.  
 XX  
 OS Clostridium botulinum.  
 XX  
 PN W0200114570-A1.  
 XX  
 XX 01-MAR-2001.  
 XX  
 XX 25-AUG-2000; 2000WO-US23427.  
 XX  
 XX 25-AUG-1999; 99US-0150710.  
 XX  
 XX (ALLR ) ALLERGAN SALES INC.  
 XX  
 XX Dolly JO, Li Y, Chan KC;  
 XX  
 XX WPI; 2001-218454/22.  
 XX  
 XX Novel isolated single-chain polypeptide derived from activatable  
 PT recombinant clostridial neurotoxin useful as therapeutic agents,  
 PT transporter molecules and adducts -  
 XX  
 XX Example 9; Page 45-47; 90pp; English.  
 XX  
 CC The present invention describes an isolated single-chain protein  
 CC comprising a therapeutic element, and a binding and translocation  
 CC element. When exposed to a protease, the therapeutic element can be  
 CC cleaved off. Examples of the binding and translocation element include  
 CC the clostridial neurotoxin H and L chains. This is useful in the  
 CC treatment of patients inoculated with the pentavalent BONT vaccine and  
 CC for delivering the therapeutic benefits of neurotoxins to patients who  
 CC are immunologically resistant to a given neurotoxin subtype, patients who  
 CC may have a lower than average concentration of receptors to a given  
 CC neurotoxin heavy chain binding group, or patients who may have a  
 CC protease-resistant variant of the membrane or vesicle toxin substrate.  
 CC The present sequence was used to demonstrate the actions of the protein  
 CC of the invention.  
 XX  
 SQ Sequence 4017 BP; 1673 A; 430 C; 565 G; 1349 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1,56e-42 Length: 4017  
 Score: 583.00 Matches: 148  
 Percent Similarity: 50.86% Conservative: 88  
 Best Local Similarity: 31.90% Mismatches: 166  
 Query Match: 24.38% Indels: 62  
 DB: 22 Gaps: 16  
 US-09-910-186a-10 (1-450) x AAF58862 (1-4017)  
 Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeAsn 21  
 Db 2692 AGTATTCCTTTTAAGCTTTCTCTTATACAGATGATAAAATTTTATTCATATTTTAT 2751  
 Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThr 41  
 Db 2752 AAATCTTTTAAAGAGATTTAAAGTACTGTTTCAGTTTAAATATGAGATATAAAATGATAAA 2811  
 Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 Db 2812 TAGCTAGATACCTTCAGGATATGATTCAAAATATAAAATATTAATGAGATGATATAAATAT 2871  
 Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 Db 2872 CCAACTAATAAAATCAATT-----GGAAATATATATGATAAACTAGTGAAGTTTAT 2925  
 Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 Db 2926 ATATCTCAAAATGATACATTATATATATATGATAATAATAATAAAATTTTAGTATTAGTTT 2985



QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117  
 |||:|||||  
 Db 2986 TGGGTAAGAAATCTCTAACTATGATTAAGATAGTAATGTTTAAATGAATGACATATA 3045  
 QY 118 IleAspSerValLys---AsnAsnSerGlyTrpSerIleGlyIleLeuSerAsnPheLeu 136  
 |||:|||||  
 Db 3046 ATAAATTCATGAGAGATAATAATTCAGGATGGAAGTATCTCTTAATCATATGAATA 3105  
 QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 |||:|||||  
 Db 3106 ATTTGGACATTCGAAGATAATTCGAGGAAATTAATCAAAATTAGCAATTTACTAGTAAAC 3165  
 QY 157 SerAsnAlaProGlyTyr---AsnLysTrpPheValThrValThrAsnAsnMet 175  
 |||:|||||  
 Db 3166 GCAATGATGTTCTGATATATAAATAGGGATTTTGTACTATACTAATGAATAGA 3225  
 QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuLeuAspThrIleLysValLysGlu 195  
 |||:|||||  
 Db 3226 TTAGGAGATTCCTAACTTTATATTAATGGAATTAATAGATCAAAATCAATTTAAAT 3285  
 QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 |||:|||||  
 Db 3286 TTAGTATATTCATGTTAGTACAAATATATTTTAAATAGTTTAAATGCTAGTTATACA 3345  
 QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 |||:|||||  
 Db 3346 ACATATATT-----GGTATTAGATATTTTAAATATTTT 3378  
 QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255  
 |||:|||||  
 Db 3379 GATAAGAAATAGTAGAAACAGAAATCAAACTTTATATAGCAATGAACTCAATACAAAT 3438  
 QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275  
 |||:|||||  
 Db 3439 ATTTTGAAGGATTTTGGGAAATTTTCTGTTTACAAAGAAATACATTTTATTAAT 3498  
 QY 276 IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn 293  
 |||:|||||  
 Db 3499 GTGTTAAACCAATAACTTTATTGATAGGAGAAAGATCTTCTTAAAGCAATTAATAAT 3558  
 QY 294 ThrArgAlaGsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310  
 |||:|||||  
 Db 3559 ATAGAAGCACTATCTTTTACGTAAATAGATATATAGTGGATATAAAGTTAAATACAA 3618  
 QY 311 ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyrPhe 329  
 |||:|||||  
 Db 3619 AGAGTAATAATAGTAGTACTAGTAAAGATAATCTGTTAGAAAGAAATGATCAGGTATATT 3678  
 QY 330 AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla 349  
 |||:|||||  
 Db 3679 AATTTTGGCCAGCAAACT-----CACATATT-----CCATTATATGCT 3720  
 QY 350 AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIle 369  
 |||:|||||  
 Db 3721 GATACAGTACCACAATAAAGAGAAACAAATAAATATCATCTGCGCAATAGATT 3780  
 QY 370 AsnAspAsnIleIlePheGluIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGln 389  
 |||:|||||  
 Db 3781 AATCAAGTAGTAGTTATG-----AATTCAGTAGGAATTTGATCAATG 3822  
 QY 390 IlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyr 409  
 |||:|||||  
 Db 3823 AATTTTAAATAAAT---AATGGAATAAT-----ATTGGTTGTTA 3861  
 QY 410 ArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsnTyrLeuValPro 425  
 |||:|||||  
 Db 3862 GTTTTCAAGGCAGATPACTGCTGCTAGTACTTGGAT----- 3900  
 QY 426 ThrValLysGlnGlyAsnTyrAlaSerLeuLeuGlnSerThrSerThrHis----- 442  
 |||:|||||  
 Db 3901 -----TATACACATATGAGATCATACACACATGATGTTT 3942  
 QY 443 TrpGlyPheVal 446

Db 3943 TGGAACTTTATT 3954  
 ||| |||:|  
 RESULT 15  
 AAV30579  
 ID AAV30579 standard; DNA; 3876 BP.  
 AC AAV30579;  
 XX  
 DT 07-DEC-1998 (first entry)  
 XX Clostridium botulinum type B toxin gene from Danish strain.  
 DE Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;  
 KW Botulism; BoTB; ds.  
 XX Clostridium botulinum serotype B Danish strain.  
 OS WO9808540-A1.  
 PN 05-MAR-1998.  
 XX 28-AUG-1997; 97WO-US15394.  
 PF 28-AUG-1996; 96US-0704159.  
 PR (OPHI-) OPHIDIAN PHARM INC.  
 PA Thalley BS, Williams JA;  
 PI WPI; 1998-230234/20.  
 DR P-PSDB; AAW68392.  
 XX Host cell containing recombinant expression vector encoding  
 Clostridium botulinum type B or E toxin - useful to treat humans  
 and other animals at risk of intoxication with clostridial toxin  
 Example 35; Page 291-296; 428pp; English.  
 CC This is the coding region of the Clostridium botulinum serotype B  
 (Danish strain) toxin gene that codes for a 1291-amino acid  
 polypeptide (see AAW68392). The C fragment (see AAW68394) of the  
 B toxin has been expressed as histidine-tagged protein in Escherichia  
 coli host cells. The invention relates to C. botulinum recombinant  
 toxin polypeptides. Methods are provided which allow for the  
 isolation of soluble recombinant proteins free of significant  
 endotoxin contamination. Preferred hosts for production of the  
 recombinant proteins are E. coli, insect cells and yeast cells.  
 CC The recombinant proteins are used as immunogens for the production  
 of vaccines and antitoxins that are useful in the treatment of  
 humans and animals at risk of intoxication with clostridial toxin.  
 XX Sequence 3876 BP; 1612 A; 370 C; 617 G; 1277 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2,26e-42 Length: 3876  
 Score: 581.00 Matches: 161  
 Percent Similarity: 49.69% Conservative: 82  
 Best Local Similarity: 32.92% Mismatches: 168  
 Query Match: 24.30% Indels: 78  
 DB: 19 Gaps: 21  
 US-09-910-186A-10 (1-450) x AAV30579 (1-3876)  
 QY 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22  
 |||:|||||  
 Db 2506 ATGCGGTTGATCTTCAATATATACCAATGATACAAATGATAGAAATGTTTAATAAA 2565  
 QY 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 |||:|||||  
 Db 2566 TATAATACGAAATTTAAATATATATCTTAATTTAGATATAGGATATATTTA 2625  
 QY 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsnPro 62



```

Db 2626 ATAGATTATCAGGATATGGCCAAAGGTAGAGGTATATGATCGAGCTGAGCTTAATGAT 2685
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 53 IlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleVal 82
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2686 -----AAAAATCAATTAAATTAACAGTACGCA-----AATAGTAGATTAGAGTC 2733
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2734 ACTCAATCAGATATCATATTAATAGTGTGTCTTCCTGATTTAGCGTAGCTTTTGG 2793
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2794 ATAAGATACCTAAATATAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2853
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2854 ATAATTAATGATGATAAATAATATCGGGTGGGAATATCTATTAGGGGTATAGGATA 2913
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2914 ATATGGACTTTAATTGATATAAATGGAATGGAACCAACCAATCGGTATTTTGAATATAACATA 2973
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 157 SerAsnAsnAlaProGlyTyr--AsnLysTrpPhePheValThrValThrAsnAsnMet 175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2974 AGAAGATATATCAGAGTATATAAATAGATAGGTGTTTTTCTACTATTACTATAATAAT 3030
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3031 TTGAATAACGCGTAAATTTATATTAAATGGAAGCTAGATAAATCAAAATCAGATATTAAGAT 3090
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3091 ATAGAGAGAGTTATGCTAATGGTGAATATATATTAAATTAGAT----- 3135
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 216 GlyLeuIleThrSerAspSerAsnIleAsn---MetTrpIleArgAspPheTyrIle 234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3136 -----GGTGATATAGATAGACACAACTTTATTTGGATGAATATTTCACTATT 3183
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3184 TTTAATACGGAATTAAGTCAATCAATATTCAAGAAAGATATAAAATTCATATATAGC 3243
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 255 AsnValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetVal 274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3244 GAATATTAAAGATTTTGGGGAATCCCTTAATGTACAAATAAGATATATATGTTT 3303
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer-----ArgGlnIle 290
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3304 AATCGGGGAATAAAATTCATATATTAACTAAAGAAAGATTCACCTGTAGTGGAATT 3363
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3364 TTAACACGTACCAATATATCAAAATCTAAATATATAATAATTATAGAGATTATATATT 3423
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 304 GlyTyrLysIleIleIleLysArgIleArgGlyAsnThr-----AsnAspThrArg 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3424 GGAGAAAAATTTATATAAAGA---AAGTCAAAATCTCAATCTATAAATGATGATATA 3480
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 321 ValArgGlyGlyAspIleLeuTyrPheAsp---MetThrIleAsnAsnLys----- 336
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3481 GTTAAAGAAAGATATATATATCTAGATTTTAAATTAATCAAGAGTCGAGAGTA 3540
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 337 ---AlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3541 TATACCTATAAATATTTAAGAAAGAGAA-----GAA 3573
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 356 AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3574 AAATGTTTGTAGCTCCTAAGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 3624
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGln---IlePheLysSerAsn 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db 3625 CAATATAAGATATCATGAACCAACACATATAGTTGTCAGTGTCTTTTAAAAA--- 3681
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 395 PheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGly 414
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3682 ---GATGAAGAAAGTACTGATGATAGAGATTGATTGGTATTCTCATCGTTTCTACGAATCT 3738
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 415 Gly-----AspTrpTyrArgHisAsn 421
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3739 GGAATTGCTATTGAGAGATATAAGATTATTTTGTATAGTAAATGGTAC----- 3789
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3790 -----TTAAAGAGGTAAAGAAACCAACCATATAATTTAAATTTG-----GGATGT 3834
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 442 HisTrpGlyPheValProValSerGlu 450
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3835 AATGGCAGTTTATTCCTTAAGATGAA 3861
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: November 7, 2002, 17:51:59  
Job time : 279 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 17:46:26 ; Search time 54 Seconds  
(without alignments)  
2555.640 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 MTIPNFSYNNLLKDI.....NYASLLESSTHWGFVPVSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh

-O=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002\_111618\_20869/app\_query.fasta\_1.647  
-DB=Issued\_Patents\_NA -QFMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MTN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09910186.ecgn.1.17 @runat\_04112002\_111618\_20869 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAEXT=7  
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586.5	24.5	3891	1 US-08-480-604A-27	Sequence 27, Appl
2	586.5	24.5	3891	2 US-08-405-496A-27	Sequence 27, Appl
3	586.5	24.5	3891	4 US-08-915-136-27	Sequence 27, Appl
4	558.5	23.4	1402	1 US-08-480-604A-25	Sequence 25, Appl
5	558.5	23.4	1402	2 US-08-405-496A-25	Sequence 25, Appl
6	558.5	23.4	1402	4 US-08-915-136-25	Sequence 25, Appl
7	557.5	23.3	1330	1 US-08-480-604A-22	Sequence 22, Appl
8	557.5	23.3	1330	2 US-08-405-496A-22	Sequence 22, Appl
9	557.5	23.3	1330	4 US-08-915-136-22	Sequence 22, Appl
10	539.5	22.6	3509	4 US-09-255-829-19	Sequence 19, Appl
11	423.5	17.7	1858	1 US-08-668-361A-6	Sequence 6, Appl
12	420	17.6	1359	1 US-07-618-312A-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-480-604A-27  
; Sequence 27, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:

13 420 17.6 1359 1 US-07-618-312A-3 Sequence 3, Appli  
14 420 17.6 1359 1 US-08-280-228-1 Sequence 1, Appli  
15 420 17.6 1359 1 US-08-280-228-3 Sequence 3, Appli  
16 402 16.8 1359 1 US-08-110-786A-7 Sequence 7, Appli  
17 155 6.5 3414 1 US-07-973-320-3 Sequence 3, Appli  
18 148.5 6.2 2934 4 US-09-206-942-52 Sequence 52, Appli  
19 148.5 6.2 2932 4 US-09-206-942-50 Sequence 50, Appli  
20 145 6.1 1430 1 US-08-276-452A-25 Sequence 25, Appli  
21 145 6.1 1430 2 US-08-798-744-25 Sequence 25, Appli  
22 142.5 6.0 1431 4 US-09-316-083-2 Sequence 2, Appli  
23 142.5 6.0 4467 1 US-08-565-907A-1 Sequence 1, Appli  
24 142.5 6.0 4467 2 US-08-910-551B-1 Sequence 1, Appli  
25 142.5 6.0 4467 2 US-08-909-425A-1 Sequence 1, Appli  
26 142.5 6.0 9636 1 US-08-323-170B-1 Sequence 1, Appli  
27 142.5 6.0 9636 4 US-08-934-441-1 Sequence 1, Appli  
28 138.5 5.8 7101 1 US-08-480-604A-9 Sequence 9, Appli  
29 138.5 5.8 7101 2 US-08-405-496A-9 Sequence 9, Appli  
30 138.5 5.8 7101 4 US-08-915-136-9 Sequence 9, Appli  
31 138.5 5.8 7101 4 US-08-957-310-9 Sequence 9, Appli  
32 132.5 5.5 1690 1 US-08-276-452A-24 Sequence 24, Appli  
33 132.5 5.5 1690 2 US-08-798-744-24 Sequence 24, Appli  
34 128 5.4 4310 3 US-09-008-172-1 Sequence 1, Appli  
35 128 5.4 4310 4 US-09-210-361-5 Sequence 5, Appli  
36 128 5.4 4310 4 US-09-740-274-5 Sequence 5, Appli  
37 126.5 5.3 10711 4 US-08-961-527-145 Sequence 145, App  
38 126 5.3 8133 1 US-08-480-604A-5 Sequence 5, Appli  
39 126 5.3 8133 2 US-08-405-496A-5 Sequence 5, Appli  
40 126 5.3 8133 4 US-08-915-136-5 Sequence 5, Appli  
41 126 5.3 8133 4 US-08-957-310-5 Sequence 5, Appli  
42 125.5 5.2 2910 4 US-09-206-942-31 Sequence 31, Appli  
43 125.5 5.2 2928 4 US-09-206-942-29 Sequence 29, Appli  
44 125.5 5.2 8920 2 US-08-446-855A-1 Sequence 1, Appli  
45 125.5 5.2 8920 4 US-09-150-741-1 Sequence 1, Appli

; APPLICATION NUMBER: US 08/329,154  
 ; FILING DATE: 25-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/161,907  
 ; FILING DATE: 02-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/985,321  
 ; FILING DATE: 04-DEC-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/429,791  
 ; FILING DATE: 31-OCT-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: INGOLIA, DIANE E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPND-01763  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 27:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3891 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..3888  
 ; US-08-480-604A-27

Alignment Scores:  
 Pred. No.: 3,46e-53 Length: 3891  
 Score: 586.50 Matches: 149  
 Percent Similarity: 48.94% Conservative: 82  
 Best Local Similarity: 31.57% Mismatches: 186  
 Query Match: 24.53% Indels: 55  
 DB: 1 Gaps: 11

US-09-910-186a-10 (1-450) x US-08-480-604A-27 (1-3891)

Qy 3 IleProPheAsnIlePheSerThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu 22  
 Db 2545 ATACCTTTTCAGCTTCCAAATACGTAGATAATCAAGATTATTATCTACATTTACTGAA 2604  
 Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 Db 2605 TATATTAGAAATATTAATACTACTCTATATTGAATTAAGATATCAAAAGTAAATCATTTA 2664  
 Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
 Db 2665 ATAGACITATCTAGGTATGCATCAAAATAAATATTTGGTAGTAAGTAATTTTGATCCA 2724  
 Qy 53 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79  
 Db 2725 ATAGATAAAATCAAAATCAATATTATTATTAAAGAGTAGT-----AAA 2769  
 Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99  
 Db 2770 ATTGAGGTAATTTAAATAATGCTATTGTAATAAATAGTATGTAATAATTTTAGTACT 2829  
 Qy 100 SerPheThrPheLeuArgIleAsnLysThrValSerAsnLeuPro-----GlyTyrThr 116  
 Db 2830 AGCTTTTGGATAAGAATTCCTAAGTATTTTAAACAGTATAAGTCTAAATGAATATACA 2889  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
 Db 2890 ATATAAATTTGATGGAAATAATATCAGATGGAAGATGATCACTTAATATGGTGAATAA 2949  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 2950 ATCTGGACTTTACAGATACCTACAGAAATAAAACAAAGACTAGTTTTTAAATACAGTCAA 3009  
 Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysThrPhePheValThrValThrAsnAsnMet 175

Db 3010 ATGATTAAATATATCAGATTATATAAACAAGATGATTTTGTAACTATCACTAATATAGA 3069  
 Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 3070 TTAATAACTCTAAATTTATATAAATGAAGATTATAAGATCAAAACCAATTTCAAT 3129  
 Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3130 TTAGGTAATATTCATGCTAGTAATAATAATGTTTAAATTTAGATGGTTGTAGAGATACA 3189  
 Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 Db 3190 -----CATAGATATATTGGATAAAATATTTTAATCTTTT 3225  
 Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnThrAsn 255  
 Db 3226 GATAAGGAATTAATGAAAGAAATCAAGATTATATGATAATCAATCAAAATCAAGT 3285  
 Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275  
 Db 3286 ATTTTAAAGACTTTTGGGTGATTATTTACAATAATGATAAACCATACTATATGTTAAAT 3345  
 Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
 Db 3346 TTATATGATCCAAATAAATAATGTCGATGTAATAATAGTATAGGTATTAGAGTTATATGAT 3405  
 Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302  
 Db 3406 CTTAAAGGGCTAGAGGTAGCGTAATGACTACAACATTTATTAAATCAAGTTTGAT 3465  
 Qy 303 GluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322  
 Db 3466 AGGGGACAAAATTTATTATAAAAAATATGCTTCGAAATAAAGATAAATATTTGTTAGA 3525  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3526 AATATATGCTGTATATATTAATGATAGTAGTTTAAATAAAGATATAGGTTA----- 3579  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3580 -----GCTACTAATGCTATCATCACAGCGAGGCGTAGAAAAATACTAAGT 3621  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3622 GCATGAATAACTGCTGATGAGGAATCTAAGTCAAGTAGTAGTATGAAGTCAAAAT 3681  
 Qy 383 ThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 3682 GATCAAGGAATAACAATAAATGCG---AAAATGATTTTACAAGATAAATAATGGGAATGAT 3738  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrPheArgHisAsnTyr 422  
 Db 3739 ATAGGCTTTATAGATTTTCATCAGTTT-----AATAAT 3771  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
 Db 3772 ATACCTAACTAGTAGCAAGTAATTTGTTATAATAGACAAATAGAAAGATCTAGTAGGACT 3831  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 3832 TTGGTTGCTCATGGGAATTTTCTCTGTAGATGAT 3867

RESULT 2  
 US-08-405-496A-27  
 ; Sequence 27, Application US/08405496A  
 ; Patent No. 5919665  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WILLIAMS, JAMES A.  
 ; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
 ; TITLE OF INVENTION: NEUROTOXIN  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:

```

; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
;
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
;
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
;
; US-08-405-496A-27
;
; Alignment Scores:
; Pred. No.: 3,46e-53 Length: 3891
; Score: 586.50 Matches: 149
; Percent Similarity: 48.94% Conservative: 82
; Best Local Similarity: 31.57% Mismatches: 186
; Query Match: 24.53% Indels: 55
; DB: 2 Gaps: 11
;
; US-09-910-186A-10 (1-450) x US-08-405-496A-27 (1-3891)
;
; Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu 22
; Db 2545 ATACCTTTTCAGCTTCCAAATACGTAGATAAATCAAGATATTATCTACATTTACTGAA 2604
;
; Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
; Db 2605 TATATAAGATATATTAATCTTCTATCTTGAATTTAAGATATGAAGTAAATCTATTA 2664
;
; Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
; Db 2665 ATAGACTTATCTAGTAGTCATCAAAAATAAATATGTTAGTAAAGTAAATTTTGATCCA 2724
;
; Qy 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
; Db 2725 ATAGATAAAAAATCAAAATTCATTAATTTAGAAAGTAGT-----AAA 2769
;
; Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99

```

```

;
; Db 2770 ATTGAGCTAAATTTTAAAAAATGCTATTGTATATAAATAGTATGTAATAATTTTAGTACT 2829
;
; Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116
; Db 2830 AGCTTTTGGATAAGAAATTCCTAAGTATTTTAAACAGTATAAGTCTAAATATGAATATACA 2889
;
; Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
; Db 2890 ATAATAATTGTAAGGAAATAATATTCAGGATGGAAGATATCACTTAATATGTTGGTAAAT 2949
;
; Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
; Db 2950 ATCTGGACTTTACAGGATACCTCAGGAAATAAACAAGAGTAGTATTTTAAATACACTCAA 3009
;
; Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
; Db 3010 ATGATTATATATCAGATTATATAACAGATGGATTTTGTAACTATCACTTAATATATA 3069
;
; Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
; Db 3070 TTAATAACTCTAAATTTATATAAATGGAAGATTAATAGATCAAAAACCAATTTCAAA 3129
;
; Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysLysLeuProAspThr 215
; Db 3130 TTAGGTAATATTCATGCTAGTAATAATATGTTAAATTTAGATCGTTGTAGAGATACA 3189
;
; Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
; Db 3190 -----CATGATATATTTGGATAAAATATTTTAATCTCTTTT 3225
;
; Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIlyThrAsn 255
; Db 3226 GATAAGGAATTAATGAAAAGAAATCAAGATTTATATGATTAATCAATCAATCAATCAGGT 3285
;
; Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
; Db 3286 ATTTAAAGACITTTGGGTGATTATTACATATGATAAACCATACTATATGTTAAAT 3345
;
; Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
; Db 3346 TTATATGATCCAAATAAATATGTCGATGTAATAATATAGTATTAGAGGTTATATGTTAT 3405
;
; Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
; Db 3406 CTTAAAGGGCTAGAGGTAGGTAAATGACTACAAACATTTATTAAATCAAGTTTGTAT 3465
;
; Qy 303 GluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
; Db 3466 AGGGGACAAAAATTTATATAAAAAATATGCTCTGGAATAAAGATAATATGTTGTTAGA 3525
;
; Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
; Db 3526 AATAATGATCGTGATATATTAATGTTAGTAAATAAATAAAGATAATAGGTTA----- 3579
;
; Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
; Db 3580 -----GCTACTAATGCATCACAGCAGCGGTAGAAAAATACTACTAAGT 3621
;
; Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
; Db 3622 GCATTAGAAATACCTGATGAGGAAATCTAAGTCAAGTAGTACTAATGAAGTCAAAAAAT 3681
;
; Qy 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
; Db 3682 GATCAAGGAATAAATAAATGTC---AAATGAATTTACAGATAATAATGGGAATGAT 3738
;
; Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyClyAspTrpTyrArgHisAsnTyr 422
; Db 3739 ATAGCTTTTATAGGATTCATCAGTTT-----AATAT 3771
;
; Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
;
;

```

Db 3772 ATAGCTAACTAGTACGAACTAATGGTATATAGACAAATAGCAAGATCTAGTAGGACT 3831

Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450

Db 3832 TTGGGTTGCTCATGGGAATTTATCTCTGTAGATGAT 3867

## RESULT 3

US-09-915-136-27  
; Sequence 27, Application US/08915136  
; Patent No. 6290560  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; ADDRESS: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/915,136  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/480,604  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01763  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3891 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3888  
; US-09-915-136-27

Alignment Scores:

Pred. No.:	3 46e-53	Length:	3891
Score:	586.50	Matches:	149
Percent Similarity:	48.94%	Conservative:	82
Best Local Similarity:	31.57%	Mismatches:	186
Query Match:	24.53%	Indels:	55
DB:	4	Gaps:	11
US-09-910-186A-10 (1-450) x US-08-915-136-27 (1-3891)			
Qy 3 IleProPheAsnIlePheSerYrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22			
Db 2545 ATACCTTTTCAGCTTCCAAATACGTAGATATCAAGATTTATTTACATTTACTGAA 2604			
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42			
Db 2605 TATATTAAGAATATTTAATCTCTATATTGAATTTAAGATATGAAGTAAATCATTTA 2664			
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62			
Db 2665 ATAGACTTATCTAGGTATGTCATCAAAATAAATATTTGGTAGTAAAGTAAATTTTGTATCCA 2724			
Qy 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79			
Db 2725 ATAGATAAAATCAATTCATTTAATTTAGAAAGTAGT-----AAA 2769			
Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetYrGluSerPheSerIle 99			
Db 2770 ATTGAGGTAATTTTAAAAAATGCTATTGTATATATAGTATGATGAAATTTTAGTACT 2829			
Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116			
Db 2830 AGCTTTGGATAGAAATTCCTAGTATTTTACAGTATTAAGTCTAAATATCAATATATA 2889			
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136			
Db 2890 ATAATAAATGTATGAAATAAATTCAGGATGAAAGTATCTACTTAATTTATGGTGAAATA 2949			
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerYrAspIle 156			
Db 2950 ATCTGGACTTTACAGGATCTCAGGAAATAAACAAGAGTAGTTTAAATACAGTCAA 3009			
Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPheValThrValThrAsnAsnMet 175			
Db 3010 ATGATTAATATATCAGATTTATATAAACAAGATGATTTTGTAACTACTACTAATAATAGA 3069			
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195			
Db 3070 TTAATAACTCTAAATTTATATAATGAAGATTAATAGATCAAAACCAATTTCAAT 3129			
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215			
Db 3130 TTAGGTAATATTCATGCTAGTAATAATAATGTTTAAATTAGATGTTGTAGAGATACA 3189			
Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235			
Db 3190 -----CATAGATATATTGGATTAATAATTTTAAATTTTAACTTTT 3225			
Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255			
Db 3226 GATAAGCAATTAATGAAGAAATCAAGATTTTATATGATAATCAATCAATCAATCAGGT 3285			
Qy 256 ValValLysAspTrpTrpGlyAsnAspLeuArgTyrAsnLysGluTyrMetValAsn 275			
Db 3286 ATTTTAAAGACTTTTGGGGTGTATTATTACAATATGATAAACCATACTATATGTTAAAT 3345			
Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287			
Db 3346 TTATATGATCCAAATAATATGTCGATGTAATAATAGTAGTATTAGAGTTTATATGTAT 3405			
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302			
Db 3406 CTTAAAGGGCGCTAGAGGTAGCGTAATGACTACAAACATTTATTTAAATTTCAAGTTGTAT 3465			

```

QY 303 GluGlyTyrLysIleIleLeuValArg 322
Db 3466 AGGGGACAAATTTATATATAAAATATGCTCTCGAATAAAGTATATATGTTAGA 3525
QY 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
Db 3526 AATAATGATCGTGTATATATATAGTAGTGTAAAAATAAAGATATAGTTA----- 3579
QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
Db 3580 -----GCTACTAATCCATCAGCAGCGGTAGAGAAAATACTAAGT 3621
QY 363 ArgGluInThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
Db 3622 GCATTAGGAATACCTGATAGGAATCTAAGTCAAGTAGTAGTAATGAAGTCAAAAAT 3681
QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 3682 GATCAGGAATAACAATAATATGTC---AAATGTAATTTACAAGATAATATGGGAATGAT 3738
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyr 422
Db 3739 ATAGGCTTTATAGGATTCATCAGTTT-----AATAAT 3771
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 3772 ATAGTAAACTAGTACGAAGTAATGGTATATATAGACAAATAGAAAGATCTAGTAGGACT 3831
QY 441 -----ThrHisTrpGlyPheValProValSerGlu 450
Db 3832 TTGGGTGCTCATGGGAATTTATTCCTGTAGATGAT 3867

RESULT 4
US-08-480-604A-25
; Sequence 25, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: FADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-480-604A-25

Alignment Scores:
Pred. No.: 8.44e-51 Length: 1402
Score: 558.50 Matches: 147
Percent Similarity: 48.73% Conservative: 83
Best Local Similarity: 31.14% Mismatches: 186
Query Match: 23.36% Indels: 56
Db: 1 Gaps: 11

US-09-910-186A-10 (1-450) x US-08-480-604A-25 (1-1402)
QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleAsnGluTyr 23
Db 45 CCATATCGAAGTCTCATATGGGTAGCATGCTCGTCTACCTTCACTGAATA 104
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 105 CATCAAGAACAATCATCAATACCTCATCTGAGCTGCTAGCAATCAATCACTGAT 164
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProil 63
Db 165 CGACCTGTCTCGTACGCTTCCAAAATCAACATCGGTCTTAAAGTTAACTTCGATCGAT 224
QY 63 ePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLysVa 80
Db 225 CGACAAGAATCAGATCCAGCTGTTCAATCTGGAATCTTCC-----AAAAT 269
QY 80 lIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe 100
Db 270 CGAAGTTATCTGGAAGATGCTATCGTATACACTCTATGATACGAAATCTCCACCTC 329
QY 100 rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrI 117
Db 330 CTTCTGGATCCGTATCCGAAATACTTCACTCCATCTCTCTGAACAATGAATACACCAT 389
QY 117 eIleAspSerValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuVa 137
Db 390 CATCAACTGCATGGAAACAATCTGTGGTGAAGATCTCTGAACTACGGTGAATCAT 449
QY 137 lPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSe 157
Db 450 CTGGACTCTGCAGGACACTCAGGAATCAACACGCGTGTGTATTCAATCACTTCAGAT 509
QY 157 rAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMe 176
Db 510 GATCAACATCTCTGACTACTCAATCGCTGGATCTCTGTTACCATCACCACCAATCGTCT 569
QY 176 tGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLe 196

```

```
Db 570 GAATAACTCCAAATCTACATCAAGCGCGTCTGATCGACCAAGAACCGATCTCCAATCT 629
Qy 196 uThrGlyLeuAsnPheserLysThrPheGluLeuAsnLysIleProAspThrGl 216
Db 630 GGGTAACATCCACCGCTCTAATACATCATGTTCAACATGGAGCGTGTGCGACACT-- 687
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetThrIleArgAspPheThrIlePheAl 236
Db 688 -----CACGCTCATCTGGATCAATCTTCAATCTGTGGA 725
Qy 236 aLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnVa 256
Db 726 CAAGAAGTACGAGAAAGAAATCAAGACCTGTACGACACCACTCCAAATCTGGTAT 785
Qy 256 lValIysAspTyrIrpGlyAsnAspLeuArgTyrAsnLysGluTyTyrMetValAsnIle 276
Db 786 CCTGAAGACTTCIGGGGTGACTACCTGCGAGTACGACCAACCGTACTACTGCTGAATCT 845
Qy 276 eAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 846 GTACGATCCGAAACAATACCTGACGTCAACAATGTAGTATCCGCGGTATACATGATCT 905
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAspPheAsnGl 303
Db 906 GAAAGGTCGCGTGTCTGTATGACTACCAACATCTACCTGAATCTCTCCGTACCG 965
Qy 303 uGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323
Db 966 TGGTACCAAAATTCATCATCAAGAAATACGCGTCTGTTAAAGCAACAATATGTTGCGAA 1025
Qy 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetIly 343
Db 1026 CAATGATCGTGTATACATCAATGTGTAGTAAAGCAACAATACCGTCTG----- 1077
Qy 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363
Db 1078 -----GCTACCAATGCTTCTCAGCGTGTGTAGAAAGATCTTGTCTGC 1121
Qy 363 gGluGlnThrLysAspIleAsnAsnIlePheGlnIleGlnProMetAsnAsnTh 383
Db 1122 TCTGAAATCCGAGCGTGTGTATCTGCTCAGGTAGTGTATGAATGAATCCAGAACGA 1181
Qy 383 rTyrTyrThrAlaSerGlnIlePheLysSerAspAsnGlyGluAsnIleSerGlyIle 403
Db 1182 CCAGGGTATCACTAACAAATGC---AAATGAATCTGCAGGACAAACAATGGTAACGATAT 1238
Qy 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyrIle 423
Db 1239 CGGTTTCATCGGTTCCACCGATC-----AACAAATAT 1271
Qy 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 1272 CGCTAAACTGGTGTCTTCCAACTGGTACAATCGTCAGATCGAAGCGTCTCTCGCACTCT 1331
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450
Db 1332 GGGTGTCTTGGGAGTTCATCCGCGTGTATGAC 1365

RESULT 5
US-08-405-496A-25
; Sequence 25, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUTROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
```

```
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLJA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-405-496A-25

Alignment Scores:
Pred. No.: 8,44e-51 Length: 1402
Score: 558.50 Matches: 147
Percent Similarity: 48.73% Conservative: 83
Best Local Similarity: 31.14% Mismatches: 186
Query Match: 23.36% Indels: 56
DB: 2 Gaps: 11

US-09-910-186A-10 (1-450) x US-08-405-496A-25 (1-1402)
Qy 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTy 23
Db 45 CCATATCGAAGTTCGTATATGCTAGCATGGCTGCTGCTGCTACCTTCACCTGAATA 104
Qy 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 105 CATCAAGACATCATCAATACCTCCATCTCTGAACTGCGCTAGCAATCCAATCCTGAT 164
Qy 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProil 63
Db 165 CGACCTGCTCTCGGTACGCTTCCAAATCAACATCGGTTCTTAAGTTAACTTCGATCCGAT 224
Qy 63 ePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLysVa 80
Db 225 CGACAGAATCATCATGCTGCTTCAATCTGGAATCTTCC-----AAAT 269
Qy 80 lIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe 100
Db 270 CGAAGTTATCTCTGAAGAAATGCTATCTATACACTCTATGTACGAAATCTTCCACCTC 329
Qy 100 rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrIle 117
Db 100 rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrIle 117
```

Db 330 CTTCTGGATCCGATATCCGGAATATCTCACTCCATCTCTCTGAAACAATGAATACACCAT 389  
QY 117 eileaspSerVallyAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhelEuva 137  
Db 390 CATCAACTGCATGGAACAATCTGTGTGTAAGAAATCTCTGAACACACGGTGAATATCAT 449  
QY 137 lPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheserTyrrAspIleSe 157  
Db 450 CTGGACTCTGCAGGACATCAGGAATCAACACAGCGTGTGTATCAAAATCTCTCAGAT 509  
QY 157 rAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValrAsnAsnMetMe 176  
Db 510 GATCAACATCTCTGACTACATCTGCTGATCTCTGTTACCAATCAACCAACATCTCT 569  
QY 176 tGlyAsnMetLysIleTyrrIleAsnGlyLysLeuIleAsnPhThrIleLysVallyGluLe 196  
Db 570 GAATACCTCAAAATCTACATCAACGCGGTCTGATGACCAAGACGATCCCAATCT 629  
QY 196 uThrGlyLeuAsnPhSerLysThrIleThrPheGluIleAsnLysIleProAspThrGl 216  
Db 630 GGTAAACATCCACGCTTCTAATACATCATCTTCAAACTGGACGTTGTCTGACACT-- 687  
QY 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAl 236  
Db 688 -----CACCGCTACATCTGATCAAAATCTCAATCTCTGATCTGTCGA 725  
QY 236 aLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrrThrAsnVa 256  
Db 726 CAAGAAGACTGACGAAAGAAATCAAGACCTCTGACGACACACCGTCCCAATCTGTGAT 785  
QY 256 lVallyAspTyrrTrpGlyAsnAspLeuArTyrrAsnLysGluTyrrMetValAsnIl 276  
Db 786 CTTGAAGACTTCTGGGTGACTACTCTGACAGCAACCGTACTACTACTGCTGATCT 845  
QY 276 eAspTyrrLeuAsnArgTyrrMetTyrrAlaAspSer----- 287  
Db 846 GTACGATCCGACAATACCTGTGCTGACGTCACAACTGATAGTATCCGCGGTACATGTACCT 905  
QY 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnGl 303  
Db 906 GAAAGGTCGCGGTGCTGTCTGATGACTACCAACATCTACCTGAACCTCTCCCTGTACCG 965  
QY 303 uGlyTyrrIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323  
Db 966 TGTGTACCAATTCATCATCAGATACAGCGCTGCTGTAACAGGACAATATCGTTCGCA 1025  
QY 323 yGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPheMetLy 343  
Db 1026 CAATGATCTGTATACATCAATGTGTGTAGTAAAGAACAAAGATACCGTCTG----- 1077  
QY 343 sAsnGluThrMetTyrrAlaAspAsnHisSerThrGluAspIleTyrrAlaIleGlyLeuAr 363  
Db 1078 -----GCTACCAATGCTCTCAGGCTGGTGTAGAAAGATCTGTGTCTGC 1121  
QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
Db 1122 TCTGGAATCCCGGAGTGTGTATCTCTCAGTGTGTATGAATCAACCAAGACGA 1181  
QY 383 rTyrrTyrrAlaSerGlnIlePheLysSerAsnPhAsnGlyLysIleSerGlyIle 403  
Db 1182 CCAGGATATCATCAAAATGCG---AAATGAATCTCAGGACCAACATGTTACAGATAT 1238  
QY 403 eCysSerIleGlyTyrrArgPheArgLeuGlyGlyAspTrpTyrrArgHisAsnTyrrLe 423  
Db 1239 CGGTTTCATCGGTTCCACCGTTC-----AACATAT 1271  
QY 423 uValProThrVallyGlnGlyAsnTyrrAlaSerLeuLeuIleSerThrSer----- 440  
Db 1272 CGCTAAACTGGTGTGCTTCCAACTGGTACAAATCGTCAGATCGAAGCTCTCTCGCACTCT 1331  
QY 441 ----ThrHisTrpGlyPheValProValSerGlu 450  
Db 1332 GGGTGTCTCTGGGAGTTCATCCCGTGTGATGAC 1365

## RESULT 6

US-08-915-136-25  
; Sequence 25, Application US/08915136  
; Patent No. 6290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/915,136  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/480,604  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1402 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1386  
US-08-915-136-25  
Alignment Scores:  
Pred. No.: 8.44e-51 Length: 1402  
Score: 558.50 Matches: 147  
Percent Similarity: 48.73% Conservative: 83  
Best Local Similarity: 31.14% Mismatches: 186  
Query Match: 23.36% Indels: 56



DB: 4 Gaps: 11  
US-09-910-186a-10 (1-450) x US-08-915-136-25 (1-1402)  
Qy 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLeuAspIleleAsnGluTyr 23  
Db 45 CCATATCGAAGGTGCTCATATGCTAGCATGGCTGCTGCTGCTGCTACCTTCACTGAAT 104  
Qy 23 rPheAsnIleAsnSerIleLeuLeuLeuGlnAsnArgIleAsnThrLeuVa 43  
Db 105 CATCAAGAACATCATCAATACCTCCATCCCTGAACCTGCGCTAGCAATCAATCACTGAT 164  
Qy 43 LaspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProI 63  
Db 165 CGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 224  
Qy 63 ePheProPheAsp-----PheIleLeuGlySerSerGlyLeuAspArgIleLeuVa 80  
Db 225 CGACAAGAACATCAGATCCAGCTGCTCAATCTGGAATCTTC-----AAAT 269  
Qy 80 lIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe 100  
Db 270 CGAAGTTATCTGAGAAATGCTATCGTATCAACTCTATGTAAGAACTTCTCCACCTC 329  
Qy 100 rPheTrpIleArgIleAsnIleValSerAsnLeuPro-----GlyTyrThrI 117  
Db 330 CTCTCGGATCGTATCCGAAATCTTCAACTCCATCTCTCTGAAACAATGAATACACCAT 389  
Qy 117 eIleAspSerValIleAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuVa 137  
Db 390 CATCAACTGCTGGAAGAAATCTGCTGGAAGAAATCTGCTGGAAGAAATCTGCTGGA 449  
Qy 137 lPheThrLeuIleAsnGlnAsnLysSerGluGlnSerIleAsnPheSerTyrAspIleSe 157  
Db 450 CTGGACTCTGAGGACACTCAGGAATCAAAACGCGTGTGTTATCAATCAATCTCTCAGAT 509  
Qy 157 rAsnAsnAlaProGlyTyr---AsnIleTyrPhePheValThrValThrAsnAsnMetMe 176  
Db 510 GATCAACATCTCTGACTACATCAATCGCTGGATCTGCTGCTGCTGCTGCTGCTGCT 569  
Qy 176 tGlyAsnMetIleTyrIleAsnGlyIleLeuIleAspThrIleIleValIleValIle 196  
Db 570 GAATAACTCCAAATCTACATCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 629  
Qy 196 uThrGlyIleAsnPheSerIleThrIlePheGluIleAsnIleIleProAspThrG 216  
Db 630 GGGTAACATCCACGCTCTTAATCAATCATGTTCAAACTGGACGGTTCGCTGACACT-- 687  
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAl 236  
Db 688 -----CACCGCTACATCTGATCAATCAATCAATCAATCAATCAATCAATCAAT 725  
Qy 236 aLysGluLeuAspGlyIleAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVa 256  
Db 726 CAAGAAGCTGAACGAAAGAAATCAAAAGACCTGTAGCAGACACCACTGCTGCTGCTG 785  
Qy 256 lValIleAspThrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnI 276  
Db 786 CTTGAAGACCTCTGGGTGACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 845  
Qy 276 eAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
Db 846 GTACGATCCGAAACAAATACGTTGAGTCAACAAATGATAGGTATCCGCGGTATACATG 905  
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnG 303  
Db 906 GAAAGGTCCGCGTGGTGTGTTAGTACTACCAACATCTACCTGAACTCTCCCTGTACCG 965  
Qy 303 uGlyTyrIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 323  
Db 966 TGGTACCAAAATTCATCATCAAGAAATACGCGTCTGCTGCTGCTGCTGCTGCTGCTG 1025  
Qy 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 343

; FILING DATE: 31-OCT-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: INGOLIA, DIANE E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPD-01763  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1330 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1314  
 ; US-08-480-604A-22

Alignment Scores:  
 Pred. No.: 1e-50  
 Score: 557.50  
 Percent Similarity: 48.80%  
 Best Local Similarity: 31.37%  
 Query Match: 23.32%  
 DB: 1  
 Gaps: 11

US-09-910-186A-10 (1-450) x US-08-480-604A-22 (1-1330)

QY 16 LeuLysAspIlelleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeu 35  
 DB 10 CTTGCTGCTACCTTCCAGTATACATCAAGAACATCAATACATCACTCCATCCCTGAACTG 69  
 QY 36 GlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 55  
 DB 70 CCTACGAAATCCATCACTGATGACCTGCTCTCGTACCTCCCAAAATCAACATCGGT 129  
 QY 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72  
 DB 130 TCTAAGTAACTTCCGATCCGATCAGCAAGAAATCAGATCCAGTGTTCATCTCGGAATCT 189  
 QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer 92  
 DB 190 TCC-----AAATCGAGTATCTCTGATCCGATCCGAAATCACTCAACTCCATC 234  
 QY 93 MetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValSerAsnLeu 112  
 DB 235 ATGTACGAAAACTTCTCCACCTCTCTGATCCGATCCGAAATCACTTCTGTTGGAAGTA 294  
 QY 113 Pro-----GlyTyrThrIleAspSerValLysAsnAsnSerGlyTrpSerIle 129  
 DB 295 TCTCTGAACATGAATACCATCATCACTGATGATGAAACAAATCTGTTGGAAAGTA 354  
 QY 130 GlyIlelleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149  
 DB 355 TCTCTGAACATGAGTAAATCATCTGCTGATCTGAGGACACTCAGGAATCAACACGCT 414  
 QY 150 IleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr---AsnLysTrpPhePhe 168  
 DB 415 GTTGATTCAATCAATCTCAGATGATCAATCACTCTGACTACATCAATCGCTGGATCTTC 474  
 QY 169 ValThrValThrAsnAsnMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle 188  
 DB 475 GTTACCATCAACCAATCGTCTGAATCACTCCAAATCTACATCAACGGCGCTGTGATC 534  
 QY 189 AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu 208  
 DB 535 GACCAAGAACCGATCTCCAATCTGGTGAACATCCACCTTCTTAATACATCATGTTCAAA 594  
 QY 209 IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIleAsnMetTyr 228  
 DB 595 CTGGACGGTTTCTGTGACACT-----CACCGCTACATCTGG 630

QY 229 IleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe 248  
 DB 631 ATCAATACCTTCAATCTGTCGCAAGAACTGAACGAAAGAAATCAAGACCTGTATAC 590  
 QY 249 AsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268  
 DB 691 GACAAACAGTCCAAATCTGATATCTGAAAGATCTTCGGGTGACTACCTGCGAGTACGAC 750  
 QY 269 LysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer--- 287  
 DB 751 AACCGTACTACATCTGTAATCTGTACGATCCGAAACAAATACGTTGACGTCAACAATGTA 810  
 QY 288 -----ArgGlnIleValPheAsnThrArg--- 295  
 DB 811 GGTATCCCGGGTTACATGTACCTGAAAGGTCGCGGTCTGTTATGACTACCAACATC 870  
 QY 296 ArgAsnAsnAsnAspPheAsnGluGlyTyrLysIlellelleLysArgIleArgGlyAsn 315  
 DB 871 TACCTGAATCTTCCCTGTACCGGTACCAATTCATCAAGAAATACGCGTCTGCT 930  
 QY 316 ThrAsnAspThrArgValArgGlyLysIlellelleLysPheAspMetThrIleAsnAsn 335  
 DB 931 AACAAAGGACAAATATCGTTCGCAACAATGATCGGTATACATCAATCTGTAGTAAAGAC 990  
 QY 336 LysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355  
 DB 991 AAGAATACCGTCTG-----GCTACCAATGCTTCTCAGGCT 1026  
 QY 356 AspIleTyrAlaIlelleGlyLeuArgGluGlnThrLysAspIleAsnAsnAsnIlellePhe 375  
 DB 1027 GGTAGAAAGATCTTGTCTGCTGGAATCCCGACGCTGTTAATCTGCTCAGGTA 1086  
 QY 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe 395  
 DB 1087 GTTGTAAATGAATCAAGACGACGAGGTATCACAAACAAATGC---AAATGAATCTG 1143  
 QY 396 AsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGly 415  
 DB 1144 CAGGACAAACATGATGATACGATCGTTTCATCGGTTCCACCACTTC----- 1191  
 QY 416 AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435  
 DB 1192 -----ACAATATCGTAAACTGTTGCTTCCAACCTGGTACATCGTCAG 1236  
 QY 436 LeuGluSerThrSer-----ThrHisTrpGlyPheValProValSerGlu 450  
 DB 1237 ATCGAAGCTTCTCTCGCACTCTGCGTGTGCTCTGGAGTTCATCCCGTGTGATGAC 1293

# RESULT 8

US-08-405-496A-22  
 ; Sequence 22, Application US/08405496A  
 ; Patent No. 5919665  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WILLIAMS, JAMES A.  
 ; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
 ; TITLE OF INVENTION: NEUROTOXIN  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MEDLEN & CARROLL, LLP  
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/405,496A  
 ; FILING DATE: 16-MAR-1995

QY	169	valthrvalthrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle	188
Db	475	GTTPACCATCACCACAAATCGTCTGTAATAACTCCAAAACTACATCAACGGCGGTCTGATC	534
QY	189	AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheLeu	208
Db	535	GACCAAGAACCGATCCCAATCTGGGTACATCCACGGTCTTAATAACATCATGTTCANA	594
QY	209	IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp	228
Db	595	CTGGACGGTGTGTCGACACT-----CACCAGTCATACCTGG	630
QY	229	IleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe	248
Db	631	ATCAAAATACCTCAATCTGTTCGACAAAGAACTGACGAAAAGAAATCAAGACCTGTGAC	690
QY	249	AsnSerLeuGlnTyrThrAsnValLysLysAspTyrTyrGlyAsnAspLeuArgTyrAsn	268
Db	691	GACACACGTCCTGATCTCTGTAAGAACTTCTGGGTGACCTACCTCAGTACGAC	754
QY	269	LysGlnTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer---	287
Db	751	AAACCGTACTACATGCTGAATCTGTACGATCCGACAAATACGTGACGTCAACAATGTA	810
QY	288	-----ArgGlnIleValPheAsnThrArg---	295
Db	811	GGTATCCGCGGTTACATGTACCTGGAAGTCGCGTGGTCTGTGTATGACTACCAACATC	870
QY	296	ArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgGlyAsn	315
Db	871	TACCTGAACCTCTCCCTGTACCGTGTGTACCAATTCATCATCAAGAAATACGCGTCTGGT	930
QY	316	ThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsn	335
Db	931	AACAAGACAATATCGTTCGCAACATGATCGTATACATCAATGTTGTAGTTAAGAAC	990
QY	336	LysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu	355
Db	991	AAAGAATACCGTCTG-----GCTACCAATGCTCTCAGGCT	1026
QY	356	AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe	375
Db	1027	GGTGTAGAAAAGATCTGTCTGTCTGGAATCCCGGACGTTGGTAATCTCTCAGGTA	1086
QY	376	GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe	395
Db	1087	GTGTATGAATGAATCCAGAACGACGAGGTATCATCAAAATGC---AAATGATCATCTG	1143
QY	396	AsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGly	415
Db	1144	CAGGACAACAATGGTAAAGATATCGGTTTCATCGGTTTCCACCAAGTTC-----	1191
QY	416	AspTyrTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu	435
Db	1192	-----AACAATATCGCTAAACTGGTGTCTTCCACTGGTACAACTCGTCAG	1236
QY	436	LeuGluSerThrSer-----ThrHisTyrGlyPheValProValSerGlu	450
Db	1237	ATGCAACGTTCTCTCGCACTCTGGGTTGCTCTTGGGAGTTTCATCCCGGTTTGATGAC	1293
RESULT 9			
US-08-915-136-22			
; Sequence 22, Application US/08915136			
; Patent No. 6290960			
; GENERAL INFORMATION:			
; APPLICANT: KINK, JOHN A.			
; APPLICANT: THALLEY, BRUCE S.			
; APPLICANT: PADHYE, NISHA V.			
; APPLICANT: FIRCA, JOSEPH R.			
; APPLICANT: STAFFORD, DOUGLAS C.			
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND			
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE			

; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MEDLEN & CARROLL, LLP  
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: US/08/915,136  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/480,504  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/405,496  
 ; FILING DATE: 16-MAR-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/329,154  
 ; FILING DATE: 25-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/161,907  
 ; FILING DATE: 02-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/985,321  
 ; FILING DATE: 04-DEC-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/429,791  
 ; FILING DATE: 31-OCT-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: INGOLIA, DIANE E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPHD-01763  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1330 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1314  
 ; US-09-915-136-22

Alignment Scores:  
 Pred. No.: 1e-50 Length: 1330  
 Score: 557.50 Matches: 144  
 Percent Similarity: 48.80% Conservative: 80  
 Best Local Similarity: 31.37% Mismatches: 180  
 Query Match: 23.32% Indels: 55  
 DB: 4 Gaps: 11

US-09-910-186a-10 (1-450) x US-08-915-136-22 (1-1330)

Db	130	TCTAAAGTTAACTTCGATCGGATCGACAGAATCAGATCCAGCTGTCAATCTGGAATCT	189
Qy	73	SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer	92
Db	190	TCC-----AAATCGAAGTTATCTCTGAGATCTATCGTATACAACTCT	234
Qy	93	MetTyrGluSerPheSerIleSerPheTyrIleArgIleAsnLysTrpValSerAsnLeu	112
Db	235	ATGTACGAAATCTCTCCAGCTCTCTGATCGGTATCCGAAATCACTTCAACTCCATC	294
Qy	113	Pro-----GlyTyrThrIleLeuAspSerValLysAsnAsnSerGlyTyrSerIle	129
Db	295	TCTCTGAACAATGAATACACATCATCACTGCATGGAACAATCTCTGTTGGAAGTA	354
Qy	130	GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer	149
Db	355	TCTCTGAACACTACGGTGAATCATCTGACTCTGCAGACACTCAGGAAATCAACACGCT	414
Qy	150	IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr--AsnLysTrpPhePhe	168
Db	415	GTTGTATTCAAAATCTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGATCTTC	474
Qy	169	ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle	188
Db	475	GTTACCATCACCAACAATCGTCTGAATACTCCAAATCTACATCAACGGCGCTCTGATC	534
Qy	189	AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu	208
Db	535	GACCAGAAACCGATCTCCAATCTGGGTAACTCCAGCTTCTAATAACATCATGTTCAA	594
Qy	209	IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp	228
Db	595	CTGGACGGTTCGTGACACT-----CACCGCTACATCTGG	630
Qy	229	IleArgaspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe	248
Db	631	ATCAAACTACTTCAATCTGTTCGACAAAGAACTGAACAAAAGAAATCAAGACCTGTAC	690
Qy	249	AsnSerLeuGlnTyrThrAsnValLysValLysAspTyrTrpGlyAsnAspLeuArgTyrAsn	268
Db	691	GACAACCAAGTCCAAATCTGGTATCTCTGAAAGACTCTCGGGGTGACTACCTCCAGTACGAC	750
Qy	269	LysGluTyrTyrMetValAsnIleAspTyrIleAsnArgTyrMetTyrAlaAsnSer---	287
Db	751	AAACCGTACTACATGCTGAATCTGTACGATCCGAAACAATACGTTGACGTCAACAATGA	810
Qy	288	-----ArgGlnIleValPheAsnThrArg---	295
Db	811	GGTATCCCGGGTTACATGACTGAAAGGTCGCGTCTGTTCTGTTATGACTACCAACATC	870
Qy	296	ArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsn	315
Db	871	TACCTGAACACTCTCCCTGTACCGTGGTACCAAAATTCATCATCAAGAAATACGCGTCTGGT	930
Qy	316	ThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsn	335
Db	931	AACAAGGACAATATCGTTCGCAACAATGATCGGTATACATCAATGTTGTAGTAGAAC	990
Qy	336	LysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu	355
Db	991	AAAGAATAACCGCTGTG-----GCTACCAATGCTCTCAGGCT	1026
Qy	356	AspIleTyrAlaIleGlyLeuArgGluThrLysAspIleAsnAspAscIleIlePhe	375
Db	1027	GGTGAGAAAAGATCTTGTCTGCTCGTCTGGAATCCCGGACGTTGTTATCTGCTCAGGTA	1086
Qy	376	GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe	395
Db	1087	GTTGTGAATGAATCAACAGACACCGAGGTATCACTAACAAATGC---AAATGAATCTG	1143
Qy	396	AsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGly	415

Qy 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72



; GENERAL INFORMATION:  
 ; APPLICANT: Brown, Robert H.  
 ; APPLICANT: Fishman, Paul S.  
 ; APPLICANT: Francis, Jonathan W.  
 ; APPLICANT: Hosler, Betsy A.  
 ; TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN  
 ; TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/668,381A  
 ; FILING DATE: 21-JUN-1996  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/000,473  
 ; FILING DATE: 23-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clark, Paul T.  
 ; REGISTRATION NUMBER: 30,164  
 ; REFERENCE/DOCKET NUMBER: 00786/269001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/542-5070  
 ; TELEFAX: 617/542-8906  
 ; TELEX: 200154  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1858 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA  
 ; US-08-668-381A-6

Alignment Scores:  
 Pred. No.: 4,08e-36 Length: 1858  
 Score: 423.50 Matches: 133  
 Percent Similarity: 44.16% Conservative: 90  
 Best Local Similarity: 26.34% Mismatches: 165  
 Query Match: 17.71% Indels: 117  
 DB: 1 Gaps: 21

US-09-910-186a-10 (1-450) x US-08-668-381A-6 (1-1858)

QY 8 PheSerThrAsnAsnSerLeuLeuLysAspIleleAsnGluTyrPheAsnAsnIle 27  
 Db 491 TTTCTTATTCTAAAAT-----CTGGATTGTTGGGTGATAATGAA 532  
 QY 28 AsnAsp-----SerLysIleLeuSerLeuGlnAsnArgLysAsn 40  
 Db 533 GAAGATATAGATGTATATATAAAGAGTACAAATTTAAATTTAGATATATAATGAT 592  
 QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60  
 Db 593 ATTATATCAGATATATCGGGTTTAATTCATCTGTAATAACATATCCAGATGCTCATTTG 652  
 QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGlu----- 75  
 Db 653 GTGCCC-----GGAATAATGCGCAAGCAATACATTTAGTA 688  
 QY 76 -----AspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMet 93  
 Db 689 RACATGATCTTCTGAAGTATATAGTCAATAAGCTATGATATATGATATATGATATG 748

QY 94 TyrGluSerPheSerIleSerPheThrIleArgIleAsnLys---TyrValSerAsnLeu 112  
 Db 749 TTTAATAATTTACCGTTAGCTTTTGGTTGAGGGTTCCTAAAGTATCTGCTGCTCAITTA 808  
 QY 113 Pro-----GlyTyrThrIleIleAspSerValLysAsnAsn----- 124  
 Db 809 GAACAATATGCGCAAAATGAGTATTCATTAATTAGCTCTATGAAACATAGCTCTCATCA 868  
 QY 125 -----SerGlyTyrPheSerIleGlyIleIleSerAspPheLeuValPheThrLeuLysGln 142  
 Db 869 ATAGGATCTGTTGGAGTGTACCTTTAAAGGTAATAACTATAATATGGACTTTAAAGAT 928  
 QY 143 AsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGly 162  
 Db 929 TCCGCGGAGAGATTAGACAAATAACTTTTAGG--GATTACCTGATAAATTAATGCT 985  
 QY 163 Tyr-----AsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180  
 Db 986 TATTAGCAATAAATGGGTTTTTATAACTATTACTAATGATAGATATTCTCTGCTAAT 1045  
 QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200  
 Db 1046 TTGTATATAATGGAGTACTTATGCGAAGTCAGAAATTACTGGTTTAGAGCTATTAGA 1105  
 QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220  
 Db 1106 GAGGATAATAATAACATTAAACTAGATAGA-----TGT 1141  
 QY 221 AspSerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240  
 Db 1142 AATATATATATCAATACGTTTCTATTGATAAATTTAGGATATTTGCCAAGCATTAAT 1201  
 QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260  
 Db 1202 CCAAAAGAGATTGAAAAATTATACACAAGTTATTTATCTATATAACCTTTTAAAGACTTC 1261  
 QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
 Db 1262 TGGGAAACCCCTTTACGATATGATACAGAAATATTTAATACCATAGCTTCTAGTTCT 1321  
 QY 277 -----AspTyrLeu-----AsnArgTyrMetTyr 284  
 Db 1322 AAAGATGTTCAATTGAAAAATATAACAGATTATATGTTTGCATAATGCGCATCGTAT 1381  
 QY 285 AlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAspPheAsnGluGly 304  
 Db 1382 ACTAACGGAAAAATGAATATATATATAGAGGTTATATAAT-----GGA 1426  
 QY 305 TyrLysIleIleLysArgIleArgGlyAsnThrAsn---AspThrArgValArgGly 323  
 Db 1427 CTAAATTTATATAAAGATATACACCTAATAATGAATAGACTCTTTTGTAAATCA 1486  
 QY 324 GlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyr----- 338  
 Db 1487 GGTGATTTTATAATTATATGTTATCATATAACAATAATGACACATTTGAGTTATCCG 1546  
 QY 339 -----AsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355  
 Db 1547 AAAGATGGAATGCTTTAATAATCTTGATAGAAATCTAAGAGTAGGTATATATGCCCA 1606  
 QY 356 AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375  
 Db 1607 GGTATCCCTCTTTATAAAAAATGGAAGCAGTAAAA----- 1642  
 QY 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe 395  
 Db 1643 -----TTGCGTGATTTAAAAACCTATCTGTACACTT-----AAATATAT 1684  
 QY 396 AsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGly 415  
 Db 1695 GATGATAAAATGATCTCT---TTAGACTAGTAGTACCCATAATGCTCAATAGGCAAC 1741

Qy 416 Asp-----TptTyrArgHisAsnTyrLeuValPro 425  
Db 1742 GATCCAAATAGGATATATTAATGCAAGCACTGGTAC---TTTAATCAATTTAAAGAT 1798  
Qy 426 ThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrHisTrpGlyPhe 445  
Db 1799 AAAATT-----TTAGGATGTGATGGTACTTT 1825  
Qy 446 ValProValSerGlu 450  
Db 1826 GTACCTACAGATGAA 1840  
RESULT 12  
US-07-618-312A-1  
; Sequence 1, Application US/07618312A  
; Patent No. 5389540  
; GENERAL INFORMATION:  
; APPLICANT: Makoff Dr, Andrew J  
; APPLICANT: Romanos Dr, Michael A  
; APPLICANT: Clare Dr, Jeffrey J  
; APPLICANT: Fairweather Dr, Neil F  
; TITLE OF INVENTION: VACCINES  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 14th Floor  
; STREET: 2200 Clarendon Boulevard,  
; CITY: Arlington,  
; STATE: Virginia,  
; COUNTRY: U.S.A.  
; ZIP: 22201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/618,312A  
; FILING DATE: 19910516  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8926832.0  
; FILING DATE: 28-NOV-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9006097.1  
; FILING DATE: 17-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford Mr, Arthur R  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 510-51  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 0101 703 8750400  
; TELEFAX: 0101 703 5253468  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1359 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEtical: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Clostridium tetani  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1356  
US-07-618-312A-1  
Alignment Scores:  
Pred. No.: 6.26e-36 Length: 1359  
Score: 420.00 Matches: 131  
Percent Similarity: 44.20% Conservative: 90  
Best Local Similarity: 26.20% Mismatches: 169

Query Match: 17.57% Indels: 110  
DB: 1 Gaps: 20  
US-09-910-186A-10 (1-450) x US-07-618-312A-1 (1-1359)  
Qy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25  
Db 7 AATCTGGATTGTGGTTCATTAATAAGAAGATATAGATTGTTATA----- 51  
Qy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45  
Db 52 ---TTAAAAAGAGTACAATTTTAATATTAGATATATAATAGATATAATATCGATATA 108  
Qy 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65  
Db 109 TCTGGGTTTAATTCATCTGTAATAACATATCCAGATGCTCAATTGGTGGCC----- 159  
Qy 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78  
Db 160 -----GGAATAAATGGCAAGCAACATACATTTAGTAAACAATGAATCTTCT 204  
Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98  
Db 205 GAAGTTATATGTCATAAGCTATGATTAATGAATATATGATATGTTTAATAATTTACC 264  
Qy 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro----- 113  
Db 265 GTTAGCTTTGGTTGAGGTTCTCTAAAGTATCTGCTAGTCATTTAGAACATATGGCACA 324  
Qy 114 ---GlyTyrThrIleIleAspSerValLysAsnAsn-----SerGlyTrp 127  
Db 325 AATGAGTATTCAATAATTAGTCTCTAGAAAAACATAGCTCTATCAATAGGATCTGGTTGG 384  
Qy 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147  
Db 385 AGTGATCACTTAAAGGTAATAACTTAATATGGACTTTAAAGATTCGCGGGAGAAATT 444  
Qy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr-----AsnLys 165  
Db 445 AGACAATAACTTTTAG---GATTACCTGATAAATTTAATGCTTATTAGCAATAAA 501  
Qy 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185  
Db 502 TGGGTTTTTATACTATTAATGATAGTATCTCTCTAAATTGTTATATAAATGA 561  
Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
Db 562 GTACTTATGGGAAGTCGCAAAATTAAGTGTAGGAGCTATTAGAGAGGATAATAATA 621  
Qy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225  
Db 622 ACATTAATAACTAGTAGA-----TGTAATAATAATAATAACA 657  
Qy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245  
Db 658 TAGCTTCTATTGATAAATTTAGGATATTTTGCAGAGCATTAATCCAAAGAGATTGAA 717  
Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu 265  
Db 718 AATTATACACAAGTTTATTTATCTATAACCTTTTAAAGAGACTTCTGGGAAACCCCTTA 777  
Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
Db 778 CGATATGATACAGAATATTATTTAATACAGCATGCTCTAGTCTTAAAGATGTTCAATGG 837  
Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289  
Db 838 AAAAATATAACAGATTATATGTTTACAAAATGCGCCCATCGTATACTAACGGAAAATTG 897  
Qy 290 IleValPheAsnThrArgArgAsnAsnAsnAsnGluGlyTyrIleIleIle 309  
Db 898 AATATATATTATAGAGGTTATATAAT-----GGACTAAATTTATTATA 942

```
QY 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuThr 328
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 943 AAAAGATATACACCTATATATGAATGAATCTTTGTTAAATCAGGTGATTTTATAA 1002
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 329 PheAspMetThrIleAsnAsnLysAlaTyR-----AsnLeu 340
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1003 TTATATGATCATATACAAATATGAGCACATCTAGGTATCCGAAAGATGGAATGCC 1062
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 341 PheMetLysAsnGluThrMetTyRAlaAspAsnHisSerThrGluAspIleTyRAlaIle 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1063 TTTAATAATCTGTAAGATTAAGAGTAGGTATTAATGCCCGCAGGTATCCCTCTTTAT 1122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1123 AAAAAATGGAAGCAGTAAAA-----TGG 1146
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 AsnAsnThrTyRThrTyRAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1147 CGTGATTAAAAACCTATTCTGTACAACTT-----AAATTATATGATGATAAAATGCA 1200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 401 SerGlyIleCysSerIleGlyThrTyRArgPheArgLeuGlyGlyAsp----- 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 TCT---TTAGGACTAGTAGTACCCATTAATGGTCAATAGGCAACGATCCAAATAGGGAT 1257
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 417 -----TTPTyRArgHisAsnTyRLeuValProThrValLysGlnGly 430
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1258 ATATTAAATGCAACCACTGGTAC---TTTAAATCAATTAAAGATAAAATT----- 1305
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 AsnTyRAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1306 -----TTAGGATGTGANTGGTACTTTGTCTACCTACAGATGAA 1341
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-07-618-312A-3
; Sequence 3, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 14th Floor
; STREET: 2200 Clarendon Boulevard,
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618,312A
; FILING DATE: 19910516
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Mr, Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
```

```
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-07-618-312A-3

Alignment Scores:
Pred. No.: 6,26e-36 Length: 1359
Score: 420.00 Matches: 131
Percent Similarity: 44.20% Conservatives: 90
Best Local Similarity: 26.20% Mismatches: 169
Query Match: 17.57% Indels: 110
DB: 1 Gaps: 20

US-09-910-186A-10 (1-450) x US-07-618-312A-3 (1-1359)
QY 6 AsnIlePheSerTyRThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyRPhaAsn 25
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7 AACCTTCTGTTGGTGCACACGAGAGACATCGATGTTATC----- 51
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 ---CTGAAAAGTCTACCATCTCAATCTGGACATCAACAACGATATATATCCGACATC 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 46 SerGlyTyRAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109 TCTGTTTCAACTCCTCTGTATCATCATATCCAGATGCTCAATGTGTGCGG----- 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 -----GGCATCAACGCGCAAGCTATCCACTGGTTAAACAACGAATCTTCT 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 LysValIleValThrGlnAsnGluAsnIleValTyRAsnSerMetTyRGlusPheSer 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 GAATTATCTGTCACAGCGCATGACATCGAATACAGACATGTTCAACACTTCACC 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro----- 113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 GTTAGCTTCTGGCTGCGCTCCGAAAGTTTCTGCTCCACCTGGACAGTACGCGACT 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 ---GlyTyRThrIleAspSerValLysAsnAso-----SerGlyTrp 127
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 AACGAGTACTCCATCATCAGCTCTATGAAGAACAACCTCCCTGCTCCATCGGCTCTGTTGG 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 385 TCTGTTTCCCTGAAGGTTAAACACCTGATCTGGACTCTGAAAGACTCCGCGGCGAAGTT 444
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 148 GlnSerIleAsnPheSerTyRAspIleSerAsnAlaProGlyTyR-----AsnLys 165
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 CGTCAGATCACTTTCCGC---GACCTGCGCGACAAGTTCAACGCGTACCTGGCTACACAA 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyRleAsnGly 185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 502 TGGGTTTTCATCACTATCAACGATCGTCTGCTCTTCTTAACCTGTACATCAACGCG 561
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 562 GTTCTGATGGCTCCGCTGAATCACTGGTCTGGGCGCTATCCGTGAGGACAAACAATC 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 622 ACTCTTAAGCTGGACCGT-----TGCAGACAAACAACACAG 657
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
Qy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspClyLysAspIleAsn 245
Db 658 TACGTATCCATCGAAGTTCCGTATCTTCTGCAAGACACTGAACCCGAAAGAGATCGAA 717
Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu 265
Db 718 AAACGTATACCACTACCTGCTATATCACTCTCTCGGTGACTTCTGGGGTAACCCGCTG 777
Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276
Db 778 CGTTACGACCGCAATATACCTGATCCCGTAGCTTCTAGTCTTAAGACGTTTCAGGTG 837
Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAACATCATCTACTACTGTTACCTGACCAACGCGCTCTACACTAACGGTAACTG 897
Qy 290 IleValPheAsnThrArgAsnAsnAspPheAsnGluGlyTyrIleIleIle 309
Db 898 AACATCTACTACCGACGCTGTACAAC-----GGCCTGAATTCATCATC 942
Qy 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328
Db 943 AAACGCTACACTCGAACAACGAATCTCTTCTGTTAAATCTGGTACCTCATCAA 1002
Qy 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340
Db 1003 CTGTACGTTTCTTACACAACACGACATCGTTGTTACCCGAAAGACGGTAACGCT 1062
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1063 TTCAACAACCTGGACAGATCTCGCTGTTGGTTACACGCTCGGGTATCCCGCTGTAC 1122
Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheClnIleGlnProMet 380
Db 1123 AAAAAAATGGAAGCTGTAAA-----CTG 1146
Qy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1147 CGTGACCTGAAACCTACTCTGTAGCTG-----AACTGTAGACGACAAAACGCT 1200
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAsp----- 416
Db 1201 TCT---CTGGGTCTGGTGTGATCCCAACGCTCAGATCGGTAAACGACCGCAACCGTGAC 1257
Qy 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430
Db 1258 ATCTGTATGCTCTACTACTGTTAC---TTCACACCTCTGAAAGACAAATC----- 1305
Qy 431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450
Db 1306 -----CTGGGTGCGACTGCTACTCTGTTCCGACCGATGAA 1341
RESULT 14
US-08-280-228-1
; Sequence 1, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanoff Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; STREET: 1100 No. 5571694th Glebe Road
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/618,312
; FILING DATE: 27-NOV-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-1
Alignment Scores:
Pred. No.: 6,26e-36 Length: 1359
Score: 420.00 Matches: 131
Percent Similarity: 44.20% Conservative: 90
Best Local Similarity: 26.20% Mismatches: 169
Query Match: 17.57% Indels: 110
DB: 1 Gaps: 20
US-09-910-186A-10 (1-450) x US-08-280-228-1 (1-1359)
Qy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25
Db 7 AATCTGGATTGTGGGTCTGTAATGAAGAGATATAGATTATATA----- 51
Qy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---TTAAAAAGAGTACAATTTTAAATTTAGATATATAATGATATTAATATCATGATATA 108
Qy 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGGGTTTAATTCATCTCTGTAATACATATCCAGATGCTCAATTCGTGCC----- 159
Qy 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78
Db 160 -----GGAATAAATGCAAGCAATACATTTAGTAACAATGAATCTTCT 204
Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATAGTCATAAAGCTATGGATATGAATATATGATATGTTTAATAATTATACC 264
Qy 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro----- 113
Db 265 GTTAGCTTTTGGTGGGTTCTCTAAAGTATCTGCTAGTCAATTAGACAATATGGCACA 324
```

QY 114 --- GlyTyrThrIleAspSerValIysAsnAsn-----SerGlyTyr 127  
|||:||||| |||:||||| :|||:||||| |||:|||||  
Db 325 AATGAGTATTCAATAATAGTCTCTATGAAACATAGTCTATCAATAGGATCTGGTTGG 384  
QY 128 SerIleGlyIleSerAspPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 385 AGTGTATCAGTAAAGGTATAACTTAATGAGCCTTAAAGATCCCGGGGAGAGTT 444  
QY 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr-----AsnLys 165  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 445 AGACAAATAACTTTTAGG--GATTACCTGATAAAATTAATGCTTATTAGCAATAAA 501  
QY 166 TrpPhePheValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 502 TGGGTTTAACTATCTACTAGTAGATTAATCTCTCTAATTTGATATAAATGGA 561  
QY 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 562 GTACTTATGGAGTGCAGAAATTAAGTCTTATAGGAGCTATTAGAGGATAATAATA 621  
QY 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIle 225  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 622 ACATTAACCTAGTAGA-----TGTAATAATAATAACAA 657  
QY 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 658 TACGTTTCTTATTGATAAATTAGGATATTTGCAAGCATTAATCCAAAGAGATTGAA 717  
QY 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValIysAspTyrTrpGlyAsnAspLeu 265  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 718 AAATTATACAAAGTATTATCTATATACCTTTTAAAGAGACTCTGGGGAACCCCTTTA 777  
QY 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 778 CGATATGATACAGATATATTATTAATCAGTAGCTTCTAGTCTTAAGATGTTCAATTG 837  
QY 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 838 AAAAATAACAGATATATATTGTTGACAAATCGCCATCGTATACTAACGGAATAATG 897  
QY 290 IleValPheAsnThrArgAsnAsnAspPheAsnGluGlyTyrLysIleIle 309  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 898 RATATATTATTAGAGGTATATAAT-----GGACTAAATTTATTATA 942  
QY 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyAspIleLeuTyr 328  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 943 AAAAGATATACACCTAATAATGAATAGATCTTTTGTAAATCAGGTGATTTTATAAA 1002  
QY 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 1003 TTATATGTATCATATACATAATAGCACATGTAGTTCATCCGAAAGATGGAATGCC 1062  
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 1063 TTTAATAATCTGTAGAAATCTAAGAGTAGTATGATTAATGCCCGGATCCCTCTTTAT 1122  
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet 380  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 1123 AAAAATAATGGAGCAGTAAAT-----TTG 1146  
QY 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 1147 CGTGATTTAAACCACTATCTGTACAACTT-----AAATTATATGATGATAAATGCA 1200  
QY 401 SerGlyIleCysSerIleGlyTyrTyrArgPheArgLeuGlyGlyAsp----- 416  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 1201 TCT---TTAGGACTAGTAGGTACCAATAGGTCAATAGGCAACGATCCAAATAGGAT 1257  
QY 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430  
|||:||||| |||:||||| |||:||||| |||:|||||  
Db 1258 ATATTATTGCAAGCACTGTAC---TTTAATCAITTAAGATAAATTT----- 1305

Qy 431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450  
Db 1306 -----TTAGGATGTGATTGGTACTTTGTACCTACAGATGAA 1341

## RESULT 15

US-08-280-228-3  
; Sequence 3, Application US/08280228  
; Patent No. 5571694  
; GENERAL INFORMATION:  
; APPLICANT: Makoff Dr, Andrew J  
; APPLICANT: Romanos Dr, Michael A  
; APPLICANT: Clare Dr, Jeffrey J  
; APPLICANT: Fairweather Dr, Neil F  
; TITLE OF INVENTION: VACCINES  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 No. 5571694th Glebe Road  
; CITY: Arlington,  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/280,228  
; FILING DATE: 25-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/618,312  
; FILING DATE: 27-NOV-1990  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8926832.0  
; FILING DATE: 28-NOV-1989  
; APPLICATION NUMBER: GB 9006097.1  
; FILING DATE: 17-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wilson, Mary J  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 117-163  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1359 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Clostridium tetani  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1356  
US-08-280-228-3

Alignment Scores:  
Pred. No.: 6.26e-36 Length: 1359  
Score: 420.00 Matches: 131  
Percent Similarity: 44.20% Conservative: 90  
Best Local Similarity: 26.20% Mismatches: 169  
Query Match: 17.57% Indels: 110  
DB: 1 Gaps: 20

US-09-910-186A-10 (1-450) x US-08-280-228-3 (1-1359)

```
Qy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlnTyrPheAsn 25
Db 7 AACCTGATTGTTGGTCGCAAGAGAGACATCGATGTTATC 51
Qy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---CTGAAAAGTCTACCATCTGCAATCGACATCAACAACGATATTATCTCCGACATC 108
Qy 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGTTTCAACTCCTCGTTTATCACAATCCAGATCTCAATIGTGCGG 159
Qy 66 PheAspPheLysLeuGlySerSerGlyGlu 78
Db 160 -----GGCATCAACGGCAAGCATATCCACCTGGTTAAACACGAATCTTCT 204
Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATCGTGCACAAGGCCATGACATCGAATACAAACAGATGTTCAACACATTCAAC 264
Qy 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro 113
Db 265 GTTACCTTCTGGCTGCGGTCGGAAGTTCTGCTCCACCTGGACAGTACGCGACT 324
Qy 114 ---GlyTyrThrIleIleAspSerValLysAsnAsn 127
Db 325 AACGAGTACTCCATCATCAGCTCTATGAAGAAACATCCCTGTCATCGGCTCTGGTTGG 384
Qy 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147
Db 385 TCTGTTCCCTGAAGGTAAACACTGATCTGGCTGACCTGAAACACTCCGCGGCGGAAGTT 444
Qy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr-----AsnLys 165
Db 445 CGTCAGATCACTTTCCCG---GACCTGCGGACAAAGTTCAACGCTACCTGGTGAACAA 501
Qy 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185
Db 502 TGGGTTTCATCACTATCAACGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
Db 562 GTTCGTATGGCTCGCGTGAATCACTGGTGGGCGCTATCGTGAGGACAAACAATC 621
Qy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIle 225
Db 622 ACTCTTAAGCTGGACCGT-----TCCACACAAACAACACAG 657
Qy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245
Db 658 TAGGTATCCATCGACAAGTTCGGTATCTTCTGCAAGCACTGAACCGGAAAGAGATCGAA 717
Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeu 265
Db 718 AAACGTATACAGCTACCTGCTATCATCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777
Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle 276
Db 778 CGTTACGACCGGATATACCTGATCCGGTACCTTCTAGCTCTAAGACGTTACGCTG 837
Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAAATCACTACCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 897
Qy 290 IleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLysIleIleIle 309
Db 898 ANCATCTACTACCGAGCTGTACAAC-----GGCTGAAATTCATCATC 942
Qy 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328
Db 943 AAACGCTACACTCCGAACACAGAAATCGATTCTTCTGTTAAATCTGTTGCTTCAACAA 1002
```

Search completed: November 7, 2002, 19:11:48

Job time : 77 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 18:38:12 ; Search time 56 Seconds

(without alignments)  
2850.955 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391

Sequence: 1 MTIFNFTSTNNSLLKDII.....NYASLLESTHWGFPVSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002.111619\_20956/app\_query.fasta\_1.647  
-DB=PublishedApplications\_NA -QFMT=fastap SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09910186.ecgn\_1\_1\_18\_rnat\_04112002.111619.20956  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2381	99.6	1371	10	US-09-350-756-3 Sequence 3, Appl1
2	2364	98.9	3876	12	US-10-051-952-4 Sequence 4, Appl1
3	863.5	36.1	3831	12	US-10-051-952-5 Sequence 5, Appl1
4	588.5	24.6	3894	12	US-10-051-952-9 Sequence 9, Appl1

## ALIGNMENTS

RESULT 1  
US-09-350-756-3  
; Sequence 3, Application US/09350756  
; Patent No. US20020034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 3  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

5	586.5	24.5	4835	10	US-09-288-326-10	Sequence 10, Appl1
6	585.5	24.5	1400	10	US-09-350-756-4	Sequence 4, Appl1
7	582	24.3	3876	12	US-10-051-952-2	Sequence 2, Appl1
8	581.5	24.3	3753	12	US-10-051-952-6	Sequence 6, Appl1
9	581.5	24.3	3759	12	US-10-051-952-7	Sequence 7, Appl1
10	578.5	24.2	1347	10	US-09-350-756-6	Sequence 6, Appl1
11	568	23.8	2452	10	US-09-350-756-7	Sequence 7, Appl1
12	561	23.5	3876	12	US-10-051-952-3	Sequence 3, Appl1
13	557.5	23.3	1338	10	US-09-350-756-1	Sequence 1, Appl1
14	553.5	23.1	3891	12	US-10-051-952-1	Sequence 1, Appl1
15	550	23.0	1327	10	US-09-350-756-9	Sequence 9, Appl1
16	546	22.8	1293	8	US-08-981-087A-5	Sequence 5, Appl1
17	546	22.8	1313	8	US-08-981-087A-6	Sequence 6, Appl1
18	541	22.6	1341	10	US-09-350-756-2	Sequence 2, Appl1
19	529	22.1	1317	10	US-09-350-756-5	Sequence 5, Appl1
20	516.5	21.6	3825	12	US-10-051-952-8	Sequence 8, Appl1
21	400.5	16.8	1987	10	US-09-350-756-8	Sequence 8, Appl1
22	393	16.4	702	10	US-09-350-756-10	Sequence 10, Appl1
23	168.5	7.0	621	10	US-09-350-756-11	Sequence 11, Appl1
24	145.5	6.1	6025	9	US-09-895-913A-11	Sequence 11, Appl1
25	142.5	6.0	1431	10	US-09-774-414-2	Sequence 2, Appl1
26	133	5.6	2796	10	US-09-815-242-4523	Sequence 4523, Ap
27	133	5.6	2799	10	US-09-815-242-8340	Sequence 8340, Ap
28	129	5.4	2523	10	US-09-815-242-4724	Sequence 4724, Ap
29	129	5.4	2526	10	US-09-815-242-8653	Sequence 8653, Ap
30	128	5.4	4310	10	US-09-740-274-5	Sequence 5, Appl1
31	121.5	5.1	15567	12	US-10-047-676A-3	Sequence 3, Appl1
32	121	5.1	1799	8	US-08-834-666A-19	Sequence 19, Appl1
33	120	5.0	640681	10	US-09-790-988-1	Sequence 1, Appl1
34	118	4.9	2979	10	US-09-815-242-4754	Sequence 4754, Ap
35	118	4.9	3009	10	US-09-815-242-8801	Sequence 8801, Ap
36	118	4.9	2376	10	US-09-815-242-9060	Sequence 9060, Ap
37	116.5	4.9	2376	10	US-09-880-107-3276	Sequence 3276, Ap
38	116.5	4.9	2988	10	US-09-815-242-4196	Sequence 4196, Ap
39	116.5	4.9	3030	10	US-09-815-242-8043	Sequence 8043, Ap
40	116.5	4.9	4389	12	US-10-044-090-611	Sequence 611, App
41	116.5	4.9	4531	10	US-09-925-300-342	Sequence 342, App
42	113.5	4.7	4594	10	US-09-070-927A-230	Sequence 230, App
43	111.5	4.7	1866	10	US-09-856-247A-1	Sequence 1, Appl1
44	111.5	4.7	2367	10	US-09-995-587A-2	Sequence 2, Appl1
45	111.5	4.7	2394	10	US-09-995-587A-3	Sequence 3, Appl1

Alignment Scores:

```

Pred. No.: 7,356-261 Length: 1371
Score: 2361.00 Matches: 448
Percent Similarity: 99.78% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 99.58% Indels: 0
DB: 10 Gaps: 0

US-09-910-186a-10 (1-450) x US-09-350-756-3 (1-1371)

Qy 1 MetThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIle 20
Db 10 ATCCACCTCCGATTCACATCTCTCTCTACACCAACAACCTCCCTGTGAAGACATCATC 69
Qy 21 AsnGluTyThrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTACTTCAACAACATCAACGACTCCAGATCTCTGCTCCCTCGAGAACCGTATGAC 129
Qy 41 ThrLeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTCGACACCTCCGCTTACACCGCGAGGTCTCCGAGGAGGTGACCTCCAGCTG 189
Qy 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAAATCTTCCCATCTCGACTTCAAGCTGGGTCTCCGCTGAGGACAGAGTAAAGTC 249
Qy 81 IleValThrGlnAsnGluAsnIleValTyThrAsnSerMetTyThrGluSerPheSerIleSer 100
Db 250 ATCGTCAACCAAGAGAGACATCGTACAACTCCATGTCAGATGTCCTTCCATCTCC 309
Qy 101 PheThrIleArgIleAsnLysTyThrValSerAsnLeuProGlyTyThrIleIleAspSer 120
Db 310 TTCTGGATCAAGATCAACAAGTGGGTCTCCAACTTGGCTGACCATCATCATCGACTCC 369
Qy 121 ValLysAsnAsnSerGlyTyThrSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGAACAACTCCGCTTGGTCTCATCGTATCATCTCCACCTTCTCGTCTTCAACCTG 429
Qy 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyThrAsnAsnAla 160
Db 430 AAGCAGACGAGGACTCGGACGATCCATCACTTCTTCCAGCATCTCCACACACGCT 489
Qy 161 ProGlyTyThrAsnLysTyThrPhePheValTyThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CCGTGTACCAACAGTGGTCTTCTCGTCCGCTCACCACCAACATGATGGTAAACATGAAG 549
Qy 181 IleTyThrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Db 550 ATCTACATCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609
Qy 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCAGACCATCACTTCCGATCAACAAGATCCACACACCGGTCTGATCACTCC 669
Qy 221 AspSerAspAsnIleAsnMetTyThrIleArgAspPheTyThrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCCGCAATATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
Qy 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyThr 260
Db 730 GGTAAAGACATCAACATCTCTTCACTCTTGGAGTACACCAACGCTGCTCAAGACTAC 789
Qy 261 TrpGlyAsnAspLeuArgTyThrAsnLysGluTyThrMetValAsnIleAspTyThrLeuAsn 280
Db 790 TGGGTACAGACCTGAGATACAAAGAGTACTTACATCTTCCGCAAGAGTGGAC 849
Qy 281 ArgTyThrMetTyThrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTACGCCCACTCCAGACAGATCTCTTCAACACACGACGATCAACACACGAC 909
Qy 301 PheAsnGluGlyTyThrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
Db 910 TTCAACGAGGGTTACAGATCATCATCAAGCGTTATCAGAGGTAAACACCAACGACACAGA 969

```

```

Qy 321 ValArgGlyGlyAspIleLeuTyThrPheAspMetThrIleAsnAsnLysAlaTyThrAsnLeu 340
Db 970 GTCAGAGGTGGTGACATCTCTGATCTTCTGACATGACTATTCACAAACAGGCTTACACCTG 1029
Qy 341 PheMetLysAsnGluThrMetTyThrAlaAspAsnHisSerThrGluAspIleTyThrAlaIle 360
Db 1030 TTCATGAAGAACGAGACCATGTACGCGCAACACCATCTCCACGAGGACATCTACGCCATC 1089
Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
Db 1090 GGTCTGGTGAGCAGACCAAGGACATCAACGACATCATCTTCCAGATCCAGCCATG 1149
Qy 381 AsnAsnThrTyThrTyThrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1150 AACCAACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAACATC 1209
Qy 401 SerGlyIleCysSerIleGlyThrTyThrArgPheArgLeuGlyGlyAspTyThrArgHis 420
Db 1210 TCCGGTATCTGTTCCATCGGTACTACAGATTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 1269
Qy 421 AsnTyThrLeuValProThrValLysGlnGlyAsnTyThrAlaSerLeuLeuSerThrSer 440
Db 1270 AACTACTTGTTCCACTGTCAAGCAGGTTAACTACGCTCTCTTGTGGAGTCCACTTCC 1329
Qy 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 1330 ACCCACTGGGGATTCGTCGCCAGTCCCGAG 1359

RESULT 2
US-10-051-952-4
; Sequence 4, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-4

Alignment Scores:
Pred. No.: 2,8e-258 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 12 Gaps: 0

US-09-910-186a-10 (1-450) x US-10-051-952-4 (1-3876)

Qy 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2527 ACAATACCTTTATATTTTTCATATACATAATTTCTTTTATTTTAAAGATATAATTAAT 2586
Qy 22 GluTyThrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2587 GAATATTTCAATAATATTAATGATTTCAAAATTTTGAGCTACAAACAGAAAAATACT 2646
Qy 42 LeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGlyAspValGlnLeuAsn 61
Db 2647 TTATGGGTATCATCGATATATGACAGAGTGGTCAAGAGCGGATGTTCACTTAAT 2706
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81

```

```
Db 2707 CCAATATTTCATTGCTTTAAATTAGGTAGTCTCAGGGAGGATAGAGTAAAGTTATA 2766
Qy 82 ValThrGlnAsnGluAsnIleValTyPheSerMetTyPheSerPheSerPhe 101
Db 2767 GTAACCCAGAAATGAATAATGATATATAATTCATGATGAAAGTTTACGATGATTTT 2826
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyPheTrpIleLeuAspSerVal 121
Db 2827 TGGATTAGATAAATAAATGGTAAGTAATTTACCTGGATATACTATAATTGATGATGTT 2886
Qy 122 LysAsnAsnSerGlyTyPheSerIleGlyIleLeuSerAsnPheLeuValPheThrLeuLys 141
Db 2887 AAAATAACTCAGGTGGAGTATAGGTATATTAGTAATTTTGTAGTATTTTACTTTAAA 2946
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyPheSerAsnAsnAlaPro 161
Db 2947 CAAATGAAGATAGTGAAACAAAGTATAAATTTAGTTATCATATATCAAAATATGCTCT 3006
Qy 162 GlyTyPheAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3007 GGATAGAATAAATGGTTTTTGTAACTGTACTAACAAATATGATGGAAATATGAAGATT 3066
Qy 182 TyPheAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3067 TATATAATGGAATTAATAGACTACTATAAAGTTTAAAGACTAACTGGAATTAATTT 3126
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3127 ACCAAAGTATAAATTTGAATAAATAAATTCAGATACCGGTTTGATTTACTTCAGAT 3186
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyPheIlePheAlaLysGluLeuAspGly 241
Db 3187 TCTGTATACATCAATATGCTGGATAGAGATTTTATATATTTGCTAAAGAAATAGATGGT 3246
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyPheAsnValValLysAspTyPhe 261
Db 3247 AAAGATATTATATATTTAAATAGCTGCAATATATACTAAATGTTGTAAGAAATATTGG 3306
Qy 262 GlyAsnAspLeuArgTyPheAsnLysGluTyPheMetValAsnIleAspTyPheLeuAsnArg 281
Db 3307 GGAATGATTTAAGATATAATAAAGAAATATATATGTTTAAATAGATTTATTAATAGA 3366
Qy 282 TyrMetTyPheAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPhe 301
Db 3367 TATATGTATGCAACTCAGCAAAATGTTTTTAATACACGTAGCAATAATAATGACTTC 3426
Qy 302 AsnGluGlyTyPheLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 3427 AATGAAGGATATAAAATTAATAAAGAAATCAGAGAAATACAAATGATAGTAGTA 3486
Qy 322 ArgGlyAspIleLeuTyPheAspMetThrIleAsnAsnLysAlaTyPheAsnLeuPhe 341
Db 3487 CGAGGAGGAGATATTATATTTTGATATGACAATTAATAACAAGCATATAATTTGTTT 3546
Qy 342 MetLysAsnGluThrMetTyPheAlaAsnHisSerThrGluAspIleTyPheAlaIleGly 361
Db 3547 ATGAAGAAATGAACATATGATGCAATATCATAGTACTGAAGATATATATGCTATAGGT 3606
Qy 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3607 TTAAGAGACAAACAAGGATATAAATGATATATATATATTTCAATACACCAATGAAT 3666
Qy 382 AsnThrTyPheTyPheAlaSerGlnIlePheLysSerAsnAsnGlyGluAsnIleSer 401
Db 3667 AATACATTATTTAGGCATCTCAAAATATTAAATCAAAATTTTAAATGAGAGAAATATTCT 3726
Qy 402 GlyIleCysSerIleGlyThrTyPheArgPheArgLeuGlyGlyAspTrpTyPheHisAsn 421
Db 3727 GGAATAGTTTCAATAGGACTTATCTGTTTTAGACTTGAGGTGATTTGGTATAGACAAAT 3786
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyPheAlaSerLeuLeuGluSerThr 441
Db 3787 TATTTGGTGCTACTGTGAAGCAAGAAATATTGTTTATTATTAAGAATCAACATCAACT 3846
```

```
Qy 442 HisTrpGlyPheValProValSerGlu 450
Db 3847 CATGGGGTTTGTACCTGTAGTGAA 3873

RESULT 3
US-10-051-952-5
; Sequence 5, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 3831
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-5

Alignment Scores:
Pred. No.: 4,71e-88 Length: 3831
Score: 863.50 Matches: 196
Percent Similarity: 61.20% Conservative: 90
Best Local Similarity: 43.46% Mismatches: 148
Query Match: 36.11% Indels: 27
DB: 11 Gaps: 11

US-09-910-186a-10 (1-450) x US-10-051-952-5 (1-3831)
```

```
Qy 2 ThrIleProPheAsnIlePheSerTyPheAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2515 ACAATGCCTTTTAAATATTTTTCATATACTAATAATTCCTTTATTAAGATATAATAAT 2574
Qy 22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2575 GAATATTTCAATAGTATTAAATGATTCACAAATTTTGAGCTTACAAACACAAAATGCT 2634
Qy 42 LeuValAspThrSerGlyTyPheAsnAlaGluValSerGluGlyAspValGlnLeuAsn 61
Db 2635 TTAGTGGATACATCAGGATATAATGCAGAGTGAGGGTAGGAGATATGTTCAACTTAAT 2694
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2695 ACGATATATACAAATGACTTTAAATTAAGTAGTTCAGAGAT-----AAATATATA 2745
Qy 82 ValThrGlnAsnGluAsnIleValTyPheAsnSerMetTyPheSerPheSerPhe 101
Db 2746 GTAAATTTTAAATAATAATATTTTATATAGCGCTATTTATGAGAACTCTAGTGTAGTTT 2805
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
Db 2806 TCGATTAAAGATATCTAAAGATTTAACTAAATTCATATGTAATGAATATCAATAATTAACGT 2865
Qy 121 ValLysAsnAsnSerGlyTyPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 2865 ATAGACAAATAATCTGGTGGAAATTTATGATAGGAATGCAATATAGATGGATTTTA 2925
Qy 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyPheSerAsnAsnAla 160
Db 2926 CAAAGATGTTTAAATAGAAAGTATAAAGTTTAAATTTTGTATTAGTGAATCATTAAGTCAT 2985
Qy 161 ProGlyTyPhe---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
Db 2986 ACAGGATATACAAATPAATAGTTTCTTCTACTAATACTAATAATAATGCGGTATATG 3045
Qy 180 LysIleTyPheAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
Db 199 LysIleTyPheAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
```

Db 3046 AACCTTTATATATGAGAAATTAAGCAGAGTCAAAAATTTGAAGATTAGATGAGGTT 3105  
 Qy 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219  
 Db 3106 AAGTTAGATAAACCATTAGTATTGGATAGATAGAGAAATATAGAT----- 3150  
 Qy 220 SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239  
 Db 3151 -----GAGAAATCAGATGCTTGGATTAGAGATTTTAATATTTTTCTTAAGAAATTA 3201  
 Qy 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnValLysAsp 259  
 Db 3202 AGTAATGAGATATTAATATGTTATGATAGGGGACAAATATTAGAAATGTTATTAAAGAT 3261  
 Qy 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeu 279  
 Db 3262 TATTGGGGAATCTTGGAGTTGATACAGATATATATATTATTAATGATATATATA 3321  
 Qy 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299  
 Db 3322 GATAGGTATATGCACCTGAAAGTAAATGTTACTTGTACTTGTGGTATCCAGATAGATCT 3381  
 Qy 300 AspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThr 319  
 Db 3382 AAATATATATCTGGAATCTTATCTATTAAATCAGTATCAGTAAGATCCTTATAGT 3441  
 Qy 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339  
 Db 3442 AGAATTTTAAATGAGGATATATAATTTCTCATATGTTATATATATAGTAGGAATATATG 3501  
 Qy 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis-----SerThrGluAsp 356  
 Db 3502 ATATAAGAGATATCTGATACATATATGACACACAGGAGGAGGTTCACAAAATGCT 3561  
 Qy 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGln 376  
 Db 3562 GTATATGCAATTAATAATACAGAGTAATTTAGTAATATATGTTAGTATAGT---ATATTAGT 3618  
 Qy 377 IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn 396  
 Db 3619 ATAAAAATATTGATCTAAATAAATAATATTGATGCAAAATTTTC---TCTAGTTTAGG 3675  
 Qy 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly 415  
 Db 3676 ---GAAATACAACTGCTCTAGCAGATATATATAAACTTGGAGATTTCTTTT----- 3726  
 Qy 416 AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435  
 Db 3727 -----AAAAATGCATAGACGCCA---GTTCCAGTAATACTAATATGAAACAAA 3771  
 Qy 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446  
 Db 3772 CTATTATCACTTCATCTTTTGGAAATTTATT 3804

## RESULT 4

US-10-051-952-9  
 ; Sequence 9, Application US/10051952  
 ; Patent No. US20020107199A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, Patricia  
 ; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
 ; FILE REFERENCE: 2933CIP  
 ; CURRENT APPLICATION NUMBER: US/10/051,952  
 ; CURRENT FILING DATE: 2002-01-17  
 ; PRIOR APPLICATION NUMBER: 09/730,237  
 ; PRIOR FILING DATE: 2000-12-05  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 9  
 ; LENGTH: 3894  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic  
 ; OTHER INFORMATION: primers used to introduce Stu I and EcoR I  
 ; OTHER INFORMATION: restriction sites into the 5' and 3' ends of the  
 ; OTHER INFORMATION: BONT/A-L chain gene fragment  
 US-10-051-952-9

Alignment Scores:  
 Pred. No.: 7.63e-57 Length: 3894  
 Score: 588.50 Matches: 155  
 Percent Similarity: 49.39% Conservative: 87  
 Best Local Similarity: 31.63% Mismatches: 171  
 Query Match: 24.61% Indels: 77  
 Db: 12 Gaps: 17

US-09-910-186a-10 (1-450) x US-10-051-952-9 (1-3894)

Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21  
 Db 2518 AGTATACCATTTGATCTTTCACATATATACCAAGCACAAATTTTAACAAAGTTTAAAT 2577  
 Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 Db 2578 AATTATATTAGTATATAGTACTGCTATTTAAAGTTTAAGTTATAGAGTGGCGGT 2637  
 Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 Db 2638 TTAATAGTATCTCTGGATATGTCGCACTATGAATGTCAGTTCAGATGTTATCTTTAAT 2697  
 Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 Db 2698 GATATAGGAATGCTCAATTTAAATTAATAATCTCT-----GAAATAGTAAATATAGC 2751  
 Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 Db 2752 GCACATCAAGTAAATTCGTTGATATGATAGTATGTTTGCATATTTTAGCATTAACITT 2811  
 Qy 102 TrpIleArgIleAsnLysTrpValSerAsn-----LeuProGlyTyr 115  
 Db 2812 TGGTAAAGGACTCTCAATAATATAATAATGATATACAACTTATCTTCAAAATGAGTAT 2871  
 Qy 116 ThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhe 135  
 Db 2872 ACAATAATTAGTGTATPAAAAATGACTCAGGATGAAAGTATCTATTAAAGGAAATAGA 2931  
 Qy 136 LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAsp 155  
 Db 2932 ATAATATGGACATTATATAGATGTTAATGCAAAATCAAAATCAATATTTTTCGAATATAGT 2991  
 Qy 156 IleSerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsn 174  
 Db 2992 ATAAAATATATATATCAGATTATATAATAATGTTTCCATAACTATTACTAATGAT 3051  
 Qy 175 MetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLys 194  
 Db 3052 AGATTAGTAACGCAAAATATTATATAATGGAAGTTTGAAAAAAGTGAATAAATTTTA 3111  
 Qy 195 GluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAsp 214  
 Db 3112 AACTAGATAGATAATTAATCTAGTAATGATATAGACTTCAAATTAATTAATGTCAGAT 3171  
 Qy 215 ThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIle 234  
 Db 3172 ACTACTAAATTT-----GTTTGGATTAAAGGATTTTAATATT 3207  
 Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254  
 Db 3208 TTTGGTAGAGAAATTAATGCTACAGAACTATCTTCATATATTTGGATTCATCATCTACA 3267  
 Qy 255 AsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrMetVal 274  
 Db 3268 AATACTTTAAAGATTTTGGGGAAATCCCTTTAAGATAGGATACACAAATACTATCTGTT 3327  
 Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThr 294

```

Db 3328 AATCAAGGTATGCAAAATATCATATATAAGTATTTTAAAGCTTCTATGGGGAACT 3387
QY 295 ArgArgAsnAsnAsnAspPheAsnGlu-----Gly 304
Db 3388 GCACACAGTACAAAC--TTTAATAATGCAGCAATAAATATCAAAATTTATATCTTGGT 3444
QY 305 TyrLysIleIleIleLysArgIle-----ArgGlyAsnThrAsnAspThrArgVal 321
Db 3445 TTACAGTATTATATAAAAAAGCATCAAAATCTCGAATATATAAATAATGATATATAGTC 3504
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3505 ACAGAAGGAGATATATATATCTTAATTT----- 3534
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-----IleTyrAla 359
Db 3535 -----GATAATATTTCTGATGAATCTTAGAGAGTATATGTT 3570
QY 360 IleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro 379
Db 3571 TTG-----GTCAATCTTAAGAAATTCAAACTCAATTAATTT-----TTAGCACCC 3615
QY 380 MetAsnAsn-----ThrTyrTyrTyrAlaSerGlnIle-----PheLysSer 393
Db 3616 ATAATGATGATGATCCTACGTTCTCTATGATCTACTACAAATAAAAAAATATTATGAAAAACA 3675
QY 394 AsnPheAsnGlyGluAsnIleSer-----GlyIleCysSerIle 406
Db 3676 ACATATAATGTGAGATACCTTGGCAAAAGATACTAAACATTTGGCGTGTGGNATT 3735
QY 407 GlyThrTyrArgPheArgLeuGlyAspTrpTyrArgHis-----AsnTyrLeuValPro 425
Db 3736 GGTAAATTTGTTAAAGATATGGATATGTTTGGGATACCTATGATATAATTTATTTT----- 3789
QY 426 ThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 3790 TGCATANGTCAGTGTATCTCAGAGAATATCTGAAATATATAAATAAATTAAGTTGGGA 3849
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 3850 TGTAATGGCAATTCATTCCTCCCGTGGATGAA 3879
```

## RESULT 5

```

US-09-288-326-10
; Sequence 10, Application US/09288326
; Patent No. US20010018049A1
; GENERAL INFORMATION:
; APPLICANT: Kei Roger Aoki
; APPLICANT: George Sachs
; TITLE OF INVENTION: Method and Compositions for the
; FILE OF INVENTION: Treatment of Pancreatitis
; FILE REFERENCE: 17282
; CURRENT APPLICATION NUMBER: US/09/288.326
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 4835
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-09-288-326-10
```

## Alignment Scores:

Pred. No.:	1,76e-56	Length:	4835
Score:	586.50	Matches:	149
Percent Similarity:	48.94%	Conservative:	82
Best Local Similarity:	31.57%	Mismatches:	186
Query Match:	24.53%	Indels:	55
DB:	10	Gaps:	11

US-09-910-186A-10 (1-450) x US-09-288-326-10 (1-4835)

```

QY 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
Db 2902 ATACCTTTTCAGCTTTCCAAATACGTAGATAATCAAGAGATTATTACTCATTTACTGAA 2961
QY 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
Db 2962 TATATTAAAGAAATATTATTAATACCTCTATATGAATTTAAGATATGAAGTAATCATTTA 3021
QY 43 ValAspThrSerGlyTyrAsnAlaGluValSerClucluglyAspValGlnLeuAsnPro 62
Db 3022 ATAGACTTATCTAGTATGTCATCAAAATATAATTTGGTAGTAAAGTAATTTTGATCCA 3081
QY 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
Db 3082 ATAGATAAAATCAAAATCAATTTAATTTAGAAAAGTAGT-----AAA 3126
QY 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99
Db 3127 ATGTGGGTAATTTTAAAAAATGCTATTGTATATAAATAGTATGATAAATTTTAGTACT 3186
QY 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116
Db 3187 AGCTTTGGATAGAATTTCTTAAGTATTTTAACAGTATAAGTCTAAATAATGAATATACA 3246
QY 117 IleLeuAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 3247 ATAATAAATTTCTATGGAATAAATTCAGGATGGAAGTATCACTTAATTTATGTTGTAATA 3306
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 3307 ATCTGGACTTTTACAGGATCTCAGGAATAAACAAGAGTAGTTTTTAATACATCACTCA 3366
QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 3367 ATGATTATATATCATCAGATTATATAACAGATGGATTTTGTGACTATCACTAATAATAGA 3426
QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 3427 TTAAATACTCTAAATTTATATAAGTGAAGATTATAGATCAAAACCAATTTCAAAT 3486
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3487 TTAGGTAATATTCATGCTAGTAAATAATATGTTTAAATAGATGTTGTAGAGATACA 3546
QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetIrpIleArgAspPheTyrIlePhe 235
Db 3547 -----CATAGATATATTGGATAAAATATTTTAATCTTTTT 3582
QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255
Db 3583 GATAAGGAATTAATGAGAAAGAAATCAAGATTTATATGATAATCAATCAAAATCAGGT 3642
QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 3643 ATTTTAAAGACTTTTGGGTGATTTTACAATATGATAAACCATCTATATGTTAAAT 3702
QY 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 3703 TTATATGATCCAAATAATATGTCGATGATAAATAATAGTAGGTATGAGGTTATATGTA 3762
QY 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
Db 3763 CTAAAGGGCCTAGAGGTAGCGTAACTACAAACATTTTATTAATCAAGTTCTGAT 3822
QY 303 GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
Db 3823 AGGGGACAAAATTTATTTATAAAAAATATGCTTCTGGAATAAAGATAATATTGTTAGA 3882
QY 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
Db 3883 AATAATGATCGGTATATATTAATGCTAGTATGTAATAAATAAAGAAATATAGCTTA----- 3936
QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
```



```

Db 3937 -----GCTACTAATGCACAGCAGGCGTAGAAAAATACTAAGT 3978
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGluIleGlnProMetAsnAsn 382
Db 3979 GCATTAGAAATACCTGATGAGGAATCTAAGTCAAGTAGTAGTAGTAAGTCAAAAAAT 4038
QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 4039 GATCAAGGAATAACAATAATATGCG--AAATGAATTTACAGATAATAATATGGGAATGAT 4095
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyr 422
Db 4096 ATAGCGTTTATAGATTTCATGAGTTT-----AATAAT 4128
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 4129 ATACCTAACTAGTAGCAAGTAATTTGGTATAATAGACAAATAGAAAGATCTAGTAGGACT 4188
QY 441 -----ThrHisTrpGlyPheValProValSerGlu 450
Db 4189 TTGGGTGCTCATGGGAATTTATTCCTGTAGATGAT 4224

```

## RESULT 6

```

US-09-350-756-4
; Sequence 4, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 4
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-4

```

## Alignment Scores:

```

Pred. No.: 3,81e-57 Length: 1400
Score: 585.50 Matches: 152
Percent Similarity: 51.72% Conservative: 89
Best Local Similarity: 32.62% Mismatches: 160
Query Match: 65 Indels: 65
DB: 10 Gaps: 17

```

US-09-910-186a-10 (1-450) x US-09-350-756-4 (1-1400)

```

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 64 AGTATCTCTTTAAGCTTCTCTTATACAGATGATATAAATTAAATTCTCTACTTCAAC 123
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 124 AAGTCTCTCAAGAGAATTAAGTCTCTTCCTGTTTAAACATGAGATACAAAGATGATAAA 183
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 184 TACGTCGACACTTCGGGTAGACTCCCAATTAACATTAACGGTGCAGGTGACAAGTAC 243

```

```

QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 244 CCAACTAACAACAAACCAATTC-----GCTATCTACAACGACAGCTTACTGAGGTGAAC 297
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 298 ATCTCTCAAAACGACATACATATCTACGACAAACAGTAGTACAAGAACTCTCTATTCTTTC 357
QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117
Db 358 TGGGTCAAGGATCTCTTAACACGACAAACAGATCGTCAACAGTTAAACAAGAGTACACATC 417
QY 118 IleAspSerValLys---AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 418 ATCAACTGATATGAGACACAACTCCGTTGGAGAGTCTCTCTTAACCAACCAACGAGATC 477
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 478 ATTTGGACCTTGAACGACACGAGGTATTAAACAAAAGTTAGCATTTCAACTPACGGTAAAC 537
QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 538 GCAACGGTATTTCTGACTACATCAACAGTGGATTTTCGTCACTATCACTAACGACAGA 597
QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 598 TTAGTGCATCTAAGCTTTTACATTAACGGTAACTTAATCGACAAAAGTCCCATTTAAAC 657
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 658 TTAGTAAACATTCACGTTCTGACAACTTTATTCAAGATCGTTAACTGCGATTACACC 717
QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
Db 718 AGATACATT-----GGCATTAGATACITCAACATTTTC 750
QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255
Db 751 GACAAGGATTTAGACGAGACGAGATTCAACTTATACAGCAACGAACTAACACCAAT 810
QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 811 ATTTTGAAGACATCTCGGGTACTACTCTGTTTACGACAAAGAACTACTACTATTAAAC 870
QY 276 IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn 293
Db 871 GTGTTAAAGCCAAACAACTTCATTGATAGGAGAAAGGATTCCTACTTTAAGCATTAACAAC 930
QY 294 ThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310
Db 931 ATCAGAAGCCTATCTTTAGCTAACAGATTATCTCTGGTATCAAGGTTAAGATCCAA 990
QY 311 ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe 329
Db 991 AGAGTTAAACAACTCTCTACTAACGATAACCTTTGTAGAAAGAACGATCAGGTCTATATT 1050
QY 330 AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla 349
Db 1051 AACTTCGTCGTAGCAAGACT-----CACTTATTC-----CCATTATATGCT 1092
QY 350 AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIle 369
Db 1093 GATACCGCTACCAACAAC-----AAGGAGAAGACCATCAAGATC 1131
QY 370 AsnAsp-----AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrAla 387
Db 1132 TCTCTCTCTGCAACAGATTTAAACCAATCGGTCTGTTATGAATCCGTCGGTAAACAACCTGT 1191
QY 388 SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly 407
Db 1192 ACCATGAATTTAAATAAT---AATGGAATAAT-----ATTGGG 1230
QY 408 ThrTyrArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsnTyrLeu 423

```

Db 1231 TGTTAGTTTCAGGCGAGACTAGTGTGCTAGTACTGGTAT 1275  
 Qy 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerHis 442  
 Db 1276 -----TATACCACATGAGAGATCACACCACAGCATGGA 1311  
 Qy 443 -----TrpGlyPheVal 446  
 Db 1312 TGTTTTGGAACTTTATT 1329

## RESULT 7

US-10-051-952-2  
 ; Sequence 2, Application US/10051952  
 ; Patent No. US20020107199A1

## GENERAL INFORMATION:

; APPLICANT: Walker, Patricia

; TITLE OF INVENTION: Methods of Administering Botulinum Toxin

; FILE REFERENCE: 2933CIP

; CURRENT APPLICATION NUMBER: US/10/051,952

; CURRENT FILING DATE: 2002-01-17

; PRIOR APPLICATION NUMBER: 09/730,237

; PRIOR FILING DATE: 2000-12-05

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 3876

; TYPE: DNA

; ORGANISM: botulinum toxin

; US-10-051-952-2

## Alignment Scores:

Pred. No.: 4,14e-56 Length: 3876  
 Score: 582.00 Matches: 161  
 Percent Similarity: 51.33% Conservative: 89  
 Best Local Similarity: 33.06% Mismatches: 163  
 Query Match: 24.34% Indels: 74  
 Gaps: 12

US-09-910-186a-10 (1-450) x US-10-051-952-2 (1-3876)

Qy 3 IleProPheAsnIlePheSerTyrThrAsnSerLeuLeuLysAspIleLeuAsnGlu 22  
 Db 2506 ATACCTTTGATCTTCACGTATCTTAATGCAATATCTAATAAATATTAATAA 2565  
 Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 Db 2566 TATAATAGCGAAATTTAAATAATATTCTTAATTTAAGATATAGAGATAATAATTA 2625  
 Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluAspValGlnLeuAsnPro 62  
 Db 2626 ATAGATTTTACAGATATGGCAAGGTAGAGTATATGATGGGTCAAGCTTAATGAT 2685  
 Qy 63 IlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleVal 82  
 Db 2686 -----AAAATCAATTTAAATTAACGTAGTTCACAGAT-----AGTAAGATTAGATC 2733  
 Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102  
 Db 2734 ACTCAAAATCAAGATATTATATTAAATAGATGATGTCCTGATTTAGCGTAGCTTTGG 2793  
 Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116  
 Db 2794 ATAAGATACCTAAATAGGAATGATGATATACAAATATATTATTCATAATGAATACG 2853  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
 Db 2854 ATAAATTAATGTATGAAATAAATTCAGCTGGAAATATCTATTAGGGGTAATAGGATA 2913  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 2914 ATATGACCTTAATGTATATAAATGGAAACCAATCAGTATTTTGTGATATACATA 2973

## RESULT 8

US-10-051-952-6

; Sequence 6, Application US/10051952

; Patent No. US20020107199A1

; GENERAL INFORMATION:

Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 Db 2974 AGAAGAGATATATACAGAGTATATAAATAGATGTTTTTTGTAATACTATTACTAATAAT 3030  
 Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 3031 TTGGATATGCTAAATTTATATTAATGGCAGCTTAGAATCAATATGATATTAAGAT 3090  
 Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3091 ATAGGAGAAGTTATTGTTAATGTTGAATTAACATTTAAATTAGAT----- 3135  
 Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsn---MetTrpIleArgAspPheTyrIle 234  
 Db 3136 -----GGTGAATGATAGACACAAATTTATTTGGATGAATATTTTACTATT 3183  
 Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254  
 Db 3184 TTTAATACGCAATTAATCAATCAATATATAAGAGAGATATATAAATTCATCATATAGC 3243  
 Qy 255 AsnValLysLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetVal 274  
 Db 3244 GAATAGTTAAAGATTTTGGGAATCCCTTTAATGTATAATAAGAAATATTATATGTT 3303  
 Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyr-----AlaAsnSerArgGlnIle 290  
 Db 3304 AATCGGGGAATAAAATTCATATATTAAACTAGTAGAAGATTCATCTAGTGAATA 3363  
 Qy 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303  
 Db 3364 TTAATACGTAGCAATATATACAGAAATCCAAATATATAAATATAGAAATTTATATAT 3423  
 Qy 304 GlyTyrLysIleIleIleLysArgIleArgGly-----AsnThrAsnAspThrArgVal 321  
 Db 3424 GGAGAAAAATTTATTATAAGAGAGAGAGCTCAATCTCAATCTATAAATGATGATAGTT 3483  
 Qy 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLys----- 336  
 Db 3484 AGAAAGAGAGATTATATACATCTAGATTTGGTACTTCACCATGAAGAGTGGAGATATAT 3543  
 Qy 337 AlaTyrAsnLeuPheMetLysAsnGlu-----ThrMetTyrAlaAspAsn 351  
 Db 3544 GCCTATAAATATTTTAAGGAACAGAGAAATATGTTTTTATCTATTATAAGTAT--- 3600  
 Qy 352 HisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAsp 371  
 Db 3601 ---TCTAATGAATTTTATAAGACTATAGAAATAAAGAA----- 3636  
 Qy 372 AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrAlaSerGlnIlePhe 391  
 Db 3637 -----TATGATGACACGCCATCA-----TATAGTTCTCAGTTGCTTTT 3675  
 Qy 392 LysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPhe 411  
 Db 3676 AAAAA-----GATGAAGAAAGTACTGATGATAGATTTGATGTTATTCATCCTTC 3729  
 Qy 412 ArgLeuGlyGly-----AspTrpTyrArgHisAsnTyrLeuVal----- 424  
 Db 3730 TAGGAATCTGAGCTTTTACGTAAAGATATAAGATTTATTTGTATAAGTAATAGTAG 3789  
 Qy 425 ---ProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerHisTrp 443  
 Db 3790 TTAAGAGAGGTAAAGAGAACCATATATAAGTCAAAATTTG-----GGATGTAATGG 3840  
 Qy 444 GlyPheValProValSerGlu 450  
 Db 3841 CAGTTTATTCCTAAAGATGAA 3861

APPLICANT: Walker, Patricia  
 TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
 FILE REFERENCE: 2933CIP  
 CURRENT APPLICATION NUMBER: US/10/051,952  
 CURRENT FILING DATE: 2002-01-17  
 PRIOR APPLICATION NUMBER: 09/730,237  
 PRIOR FILING DATE: 2000-12-05  
 NUMBER OF SEQ ID NOS: 12  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 3753  
 TYPE: DNA  
 ORGANISM: botulinum toxin  
 US-10-051-952-6

Alignment Scores:  
 Pred. No.: 4.5e-56 Length: 3753  
 Score: 581.50 Matches: 153  
 Percent Similarity: 52.14% Conservative: 91  
 Best Local Similarity: 32.69% Mismatches: 153  
 Query Match: 24.32% Indels: 71  
 DB: 12 Gaps: 19

US-09-910-186A-10 (1-450) x US-10-051-952-6 (1-3753)

QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
 Db 2464 AGTATCTCTTTAGCTTCTCTTCTTATACAGATGATAAATTTTATTCATATTTAA 2523  
 QY 22 GlutyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 Db 2524 AAGTCTTTAAGAGATTAAAGTAGTCTGTTTAAATATGAGATATAAATAATGATAA 2583  
 QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 Db 2584 TAGTAGATACCTCAGGATATGATCAATATAAATAATTAATGAGATATATAAATAT 2643  
 QY 62 ProilePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValle 81  
 Db 2644 CCAACTAATAAAATCAATTT-----GGAATATATAATGATAAACTTAGTGAAGTTAA 2697  
 QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 Db 2698 ATATCTCAAAATGATTACATATATATATGATAAATAAATTTTATGATTTAGTTT 2757  
 QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117  
 Db 2758 TGGGTAGAATTCCTAATGATAAATAGATAGTAATGTTAATGATGATGATGATA 2817  
 QY 118 IleAspSerValLys---AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
 Db 2818 ATAAATTTGATGAGGATAAATATTCAGGATGGAAGATATCTCTTAATCATATGAATA 2877  
 QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 2878 ATTTGGACATTCGAGATAATTCAGGATTAATCAAAATACATTAATGATGTAAC 2937  
 QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 Db 2938 GCAAAATGTTATTTCTGATTATATAAATAAAGTGGATTTTCTGAATATAATGATAGA 2997  
 QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 2998 TTAGGAGATTTCAACTTATATTATTAATGGAATTTAATAGATAAATAAATCAATTTAA 3057  
 QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3058 TTAGTATATATTCAGTTAGTGACAAATATATTTTAAATAGTTAAATTTAGTTATACA 3117  
 QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 Db 3118 AGATATATTT-----GGTATTAGATATTTTAAATTTT 3150

QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAsn 255  
 Db 3151 GATAAAGAATTAGATGAACAGAAATCAAACTTTATATAACAATGAACCAATGCAAAAT 3210  
 QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275  
 Db 3211 ATTTTAAGGATTTTGGGAAATATTGCTTTATGACAAAGATAGTATTATTAAT 3270  
 QY 276 Ile-----AspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleVal 291  
 Db 3271 GTGTTAAACCAAAATAACTTTATTAATAGG-----AGAACAGATCTACATTAAGCAAT 3324  
 QY 292 PheAsnThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIle 308  
 Db 3325 AATATATATAGAGACACTATTTTATAGCTATATAGTATATAGTGAATAAAAGTTAAA 3384  
 QY 309 IleLysArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeu 327  
 Db 3385 ATACAAAGATTAATAATAGTAGTACTACAGATAATCTGTTTAGAAAGATGATCAGGTA 3444  
 QY 328 TyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMet 347  
 Db 3445 TATATTAATTT---GTACGACGACAACTCACTTACTT-----CCATTA 3486  
 QY 348 TyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLys 367  
 Db 3487 TAGCTGATACAGCTACCAAAAT-----AAAGAGAAACAAATA 3525  
 QY 368 AspIleAsnAsp-----AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyr 385  
 Db 3526 AAAATATCATCATCTGGCAATAGATTAAATCAAGTAGTAGTATGATTAATTCAGTA----- 3579  
 QY 386 TyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSer 405  
 Db 3580 GGATGTCATCAATGAATTTTAAATAAT---AATGGAATAAT----- 3618  
 QY 406 IleGlyThrTyrArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsn 421  
 Db 3619 ATTTGGGTGTAGTTTCAAGCAGACTAGTGTGTAGTGTAGTACTTGGTAT----- 3669  
 QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
 Db 3670 -----TATACACATATGAGAGATAATAACAAACAGC 3699

RESULT 9  
 US-10-051-952-7  
 ; Sequence 7, Application US/10051952  
 ; Patent No. US20020107199A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, Patricia  
 ; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
 ; FILE REFERENCE: 2933CIP  
 ; CURRENT APPLICATION NUMBER: US/10/051,952  
 ; CURRENT FILING DATE: 2002-01-17  
 ; PRIOR APPLICATION NUMBER: 09/730,237  
 ; PRIOR FILING DATE: 2000-12-05  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 3759  
 ; TYPE: DNA  
 ; ORGANISM: botulinum toxin  
 ; US-10-051-952-7

Alignment Scores:  
 Pred. No.: 4.52e-56 Length: 3759  
 Score: 581.50 Matches: 152  
 Percent Similarity: 51.72% Conservative: 89  
 Best Local Similarity: 32.62% Mismatches: 160

Query Match:	24.32%	Indels:	65
DB:	12	Gaps:	17
US-09-910-186A-10 (1-450) x US-10-051-952-7 (1-3759)			
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuGlnAspIleLeuAsn 21			
DB 2464 AGTATCCCTTTAAGCTTCTCTTATACAGATGATAAAATTTTAATTCATATTTAAT 2523			
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41			
DB 2524 AAATCTTTTAAAGAGAAATAAAAGTAGTTCACTTTTTAAATATGACATATAAAATGATAAA 2583			
QY 42 LeuValasThrSerGlyTyrAsnLagLluValSerGluGluGlyAspValGlnLeuAsn 61			
DB 2584 TAGCTAGATACCTTCAGGATATGATGCTCAATATAATATAATTAATGAGATGCTATATAATAT 2643			
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81			
DB 2644 CCAACTAAATAAAATCAATTT-----GGAATATATAATGATAAATCTAGTGAAGTTAAT 2697			
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101			
DB 2698 ATATCTCAAAATGATAGATTATATATGATATAATAATAATAAAATTTTGTATTAGTTAGTTT 2757			
QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117			
DB 2758 TGGGTAGAAATTCCTAACCTATGATATAAGATAGTAAATGTTAATAATGAATAGACTATA 2817			
QY 118 IleAspSerValLys---AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136			
DB 2818 ATAAATGTGATGAGAGATTAATATTCAGATGGAAAGTATCTCTTATCATATGAATGAATA 2877			
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156			
DB 2878 ATTGTGCACATGCAAGATTAATCAGGAATTAATCAAAATTAGCATTTAAGTATGTTAAC 2937			
QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175			
DB 2938 GCAATGGTATTTCTGATTATATAATATAGTGAGATTTTGTGAATCACTAACCTATGATAGA 2997			
QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195			
DB 2998 TTAGGACATCTAACTTTATATTAATGAAATTTAATAGATCAAAATCAATTTTAAAT 3057			
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215			
DB 3058 TTAGTAAATATTCACTTGTAGTACATATATTATTTAAATAGTTTAATGTAGTTATACA 3117			
QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235			
DB 3118 AGATATATT-----GGTATTAGATATTTTAAATATTTT 3150			
QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255			
DB 3151 GATAAAGAAATTAGTGAACAGAAATTCAACTTTATATGACATGAACCTTAATACAAAT 3210			
QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275			
DB 3211 ATTTTGAAGCATTTTGGGAAATTTTTCCTTTATGACAAAGATACATATTATTAAAT 3270			
QY 276 IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn 293			
DB 3271 GTGTATAAACCAATAACTTTTATGTATGAGGAAAGATTTCTACCTTAAGCATTAATAAT 3330			
QY 294 ThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleIleLys 310			
DB 3331 ATAGAGACCACTATCTTTTAGCTATAGATTATATAGTGGAAATAAAGTTAAATACAA 3390			
QY 311 ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe 329			
DB 3391 AGAGTTAATATAGTACTACAGATAATCTTGTTAGAAAGAAATGATCAGGTATATATT 3450			

Qy 330 AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla 349  
 Db 3451 AATTTGGTAGCCAGAAACT-----CACTATTT-----CCATTAAGCT 3492  
 Qy 350 AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluInThrLysAspIle 369  
 Db 3493 GATACAGCTACCACAAT-----AAAGAGAAAAACAATAAAAAATA 3531  
 Qy 370 AsnAsp-----AsnIleIlePheGlnIleGlnPromMetAsnAsnThrTyrTyrAla 387  
 Db 3532 TCATCATCTGGCAATAGAGTTAATCAGTAGTAGTAGTTAAGTAATTCAGTAGAATAATGCT 3591  
 Qy 388 SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyLeuCysSerIleGly 407  
 Db 3592 ACAATGAATTTAAAAAATAAT--AATGGAAATAAT-----ATTGGG 3630  
 Qy 408 ThrTyrArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsnTyrLeu 423  
 Db 3631 TTGTTAGGTTTCAGGCAGACACTAGTCTGCTAGTACTTGGTAT----- 3675  
 Qy 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHis--- 442  
 Db 3676 -----TATACATATGAGAGATCATCAACACGCAATGGA 3711  
 Qy 443 -----TrpGlyPheVal 446  
 Db 3712 TGTTTYGGAACTTTATT 3729  
 RESULT 10  
 US-09-350-756-6  
 ; Sequence 6, Application US/09350756  
 ; Patent No. US20020034521A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
 ; APPLICANT: John S. Lee  
 ; APPLICANT: Peter Pushko  
 ; APPLICANT: Michael D. Parker  
 ; APPLICANT: Jonathan F. Smith  
 ; APPLICANT: Mark T. Dertzbaugh  
 ; APPLICANT: Leonard Smith  
 ; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
 ; FILE REFERENCE: 003/124/SAP RIID 98-21  
 ; CURRENT APPLICATION NUMBER: US/09/350,756  
 ; CURRENT FILING DATE: 1999-07-09  
 ; EARLIER APPLICATION NUMBER: US 60/092,416  
 ; EARLIER FILING DATE: 1998-07-10  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
 ; SEQ ID NO 6  
 ; LENGTH: 1347  
 ; TYPE: DNA  
 ; ORGANISM: Artificial  
 ; FEATURE:  
 ; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
 US-09-350-756-6  
 Alignment Scores:  
 Pred No.: 2,24e-56 Length: 1347  
 Score: 578.50 Matches: 151  
 Percent Similarity: 49.27% Conservative: 84  
 Best Local Similarity: 31.68% Mismatches: 169  
 Query Match: 24.19% Indels: 73  
 DB: 10 Gaps: 16  
 US-09-910-186A-10 (1-450) x US-09-350-756-6 (1-1347)  
 Qy 13 AsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIle 32  
 Db 7 GACACCACTCTGATCCAGGTCCTCAACAACTACTCTCCACACTCTCTCTCCACGCCATC 66  
 Qy 33 LeuSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluVal 52  
 Db 67 CTGTCCTGTCCTACCGTGTGCTGCTGATGCTGACTCTCTCGGTTTACGAGCGACCATC 126

```
QY 53 SerGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGlySer 72
   ::::::::::::::::::::
Db 127 AACGTCGGTTCGACGCTCATCTTCAACGACATCGGTACGGTCAAGCTGAACAC 186
   ::::::::::::::::::::
QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGlnValValThrAsnSer 92
   ::::::::::::::::::::
Db 187 TCC-----GAGAACTCCAACATCAGCCGCCACCGTCCAGTTCGTCTACGACTCC 240
   ::::::::::::::::::::
QY 93 MetThrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValSerAsn--- 111
   ::::::::::::::::::::
Db 241 ATGTTGACAACTTCCATCAACTTCTGGTCCGTATCCCAAGTACAAACACACACAC 300
   ::::::::::::::::::::
QY 112 -----LeuProGlyThrThrIleLeuAspSerValIleAsnAsnSerGly 126
   ::::::::::::::::::::
Db 301 ATCCAGACCTACCTGACAGACGAGTACCATCATCTCTGTATCAAGAACGACTCCCGT 360
   ::::::::::::::::::::
QY 127 TrpSerIleGlyIleLeuSerAsnLeuValPheThrLeuLysGlnAsnGluAspSer 146
   ::::::::::::::::::::
Db 361 TGAAGGTCCTCCATCAAGGAAACCGTATCATCTGGACCTGATCGACGTCAACGCCAAG 420
   ::::::::::::::::::::
QY 147 GluGlnSerIleAsnPheSerThrAspIleSerAsnAlaProGlyThr---AsnLys 165
   ::::::::::::::::::::
Db 421 TCCAGTCCATCTCTTCGAGTACTCCATCAGGACACATCTCCGACTACATCAACAG 480
   ::::::::::::::::::::
QY 156 TrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTrpIleAsnGly 185
   ::::::::::::::::::::
Db 481 TGGTTCCTCATCACCACCAACGACGCTGGTGAACGCCAACATCATCATCAACGGT 540
   ::::::::::::::::::::
QY 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
   ::::::::::::::::::::
Db 541 TCCCTGGAAGAGTCCGAGAGATCCGTGACCTGGACCGTATCAACCTCCCAACGACATC 600
   ::::::::::::::::::::
QY 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225
   ::::::::::::::::::::
Db 601 GACTCAAGCTGAGTACTGACGACACCAACCAAGTTC----- 639
   ::::::::::::::::::::
QY 226 AsnMetTrpIleArgAspPheTrpIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245
   ::::::::::::::::::::
Db 640 ---GTCTGGATCAAGGACTTCAACATCTTCGTCGGTGAACCGCCACCGAGGTCTCC 696
   ::::::::::::::::::::
QY 246 IleLeuPheAsnSerLeuGlnThrValValLysValLysAspTrpTrpGlyAsnAspLeu 265
   ::::::::::::::::::::
Db 697 TCCCTGTACTGATCCAGTCTCCACCAACACCCCTGAGGACTTCTGGGAAACCCACTG 756
   ::::::::::::::::::::
QY 266 ArgTrpAsnLysGluThrValAsnIleAspTrpLeuAsnArgTrpMetTrpAla 285
   ::::::::::::::::::::
Db 757 CGTACGACACCCAGTACTACCTGTTCAACCCAGGATGACGAAACATCTACATCAAGTAC 816
   ::::::::::::::::::::
QY 286 AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGlu----- 303
   ::::::::::::::::::::
Db 817 TCTCCAGGCTCCATGGGTGAGACCGCCCTCGTACCAAC---TTCACACACGCCGCC 873
   ::::::::::::::::::::
QY 304 -----GlyTrpLysIleIleIleLysArgIle----- 312
   ::::::::::::::::::::
Db 874 ATCACTACCAAGAACCTGTACTGTCGTCGTTTCATCATCAAGAACGCCCTCCAACTCC 933
   ::::::::::::::::::::
QY 313 ArgGlyAsnThrAsnAspThrArgValArgGlyLysAspIleLeuThrPheAspMetThr 332
   ::::::::::::::::::::
Db 934 CGTAACATCAACACGACACATCGTCCGTGAGGCTGACTACATCTACCTGACATC--- 990
   ::::::::::::::::::::
QY 333 IleAsnAsnLysAlaTyAsnLeuPheMetLysAsnGluThrMetTrpAlaAspAsnHis 352
   ::::::::::::::::::::
Db 991 -----GACACATC 999
   ::::::::::::::::::::
QY 353 SerThrGluAspIleValAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn 372
   ::::::::::::::::::::
Db 1000 TCCGACGAGTCC---TACCGTGTCTACGTCCTGTCACCTCCCAAGGAGATCCAGCCACG 1056
   ::::::::::::::::::::
QY 373 IleIlePheGlnIleGlnProMetAsnAsn-----ThrTrpTrpTrpAlaSerGlnIle 390
   ::::::::::::::::::::
Db 1057 CTGTTC-----CTGGCCCAATCAACGACGACCCCTACCTTCTACGACGCTCTGCAGATC 1110
   ::::::::::::::::::::
QY 391 -----PheLysSerAsnPheAsnGlyGluAsnIleSer----- 401
   ::::::::::::::::::::
Db 1111 AGAAGTACTAGAGAGACACACCTACACTCTCAGATCTCAGATCTCGGAGAGACACCAAG 1170
   ::::::::::::::::::::
QY 402 -----GlyIleCysSerIleGlyThrTrpArgPheArgLeuGlyGlyAspTrpTrpArg 419
   ::::::::::::::::::::
Db 1171 ACCTTCGACTGTTCGGTATCGTAAAGTTCGTCAGGACTACGGTTCAGTCTGGGACACC 1230
   ::::::::::::::::::::
QY 420 His---AsnThrLeuValProThrValLysGlnGlyAsnTrpAlaSerLeuLeuGluSer 438
   ::::::::::::::::::::
Db 1231 TAGACAACACTACTTC-----TGTATCTCCAGTGGTACCTCGCTCGTATCCGAGAAC 1284
   ::::::::::::::::::::
QY 439 ThrSer-----ThrHisTrpGlyPheValProValSerGlu 450
   ::::::::::::::::::::
Db 1285 ATCAACAAGCTGGTCTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
   ::::::::::::::::::::
RESULT 11
US-09-350-756-7
; Sequence 7, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 7
; LENGTH: 2452
; TYPE: DNA
; ORGANISM: Clostridium botulinum
; FEATURE:
;
US-09-350-756-7
Alignment Scores:
Pred. No.: 8-28e-55 Length: 2452
Score: 568.00 Matches: 149
Percent Similarity: 49.04% Conservative: 82
Best Local Similarity: 31.63% Mismatches: 184
Query Match: 23.76% Indels: 57
DB: 10 Gaps: 11
US-09-910-186A-10 (1-450) x US-09-350-756-7 (1-2452)
QY 3 IleProPheAsnIlePheSerTrpThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
   ::::::::::::::::::::
Db 1116 ATACCTTTTCAGCTTTCCTCAAAATACGATGATATCAAGATTTATTATCTACATTCTGAA 1175
   ::::::::::::::::::::
QY 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
   ::::::::::::::::::::
Db 1176 TATATTAAAGATATTATTAATCTCTATATTGAATTAAGATATCAAAAGTAAATCATTTA 1235
   ::::::::::::::::::::
QY 43 ValAspThrSerGlyTrpAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
   ::::::::::::::::::::
Db 1236 ATAGACTTATCTAGGTATGATCAATAAATAATATTGGTAGTAAGATTAATTTGATCCA 1295
   ::::::::::::::::::::
QY 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
   ::::::::::::::::::::
Db 1296 ATAGATAAAATCAAAATTCATTTATTTATTTAGAAAGTAGT-----AAA 1340
   ::::::::::::::::::::
QY 80 ValIleValThrGlnAsnGluAsnIleValTrpAsnSerMetTrpGluSerPheSerIle 99
   ::::::::::::::::::::
Db 1341 ATTGAGGTAATTTTAAATAATGCTATTGTTATATTAATAGTATGATGATAAATTTTACT 1400
   ::::::::::::::::::::
```

Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
Db 1401 AGCTTTTGGATAAGAACTCTAAGTATTTTACAGTATAAGCTAAATAATGAATATACA 1460  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 1461 ATAATAAGTATGGAATAATATCAGGATGGAAGATCATCTTAATATGGTGAATA 1520  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
Db 1521 ATCTGGACTTACAGGATACCTCAGGAATAAACAAGAGTAGTTTAAATAC----- 1574  
Qy 157 SerAsnAsnAlaProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMet 176  
Db 1575 AGTCA-AATGATTAATATACACAGATGGATTTTGTAACATCATCTAATAATAGATTA 1633  
Qy 177 GlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeu 196  
Db 1634 AATAACTCTAAATTTATATAATGGAAGATTAATAGATCAAAACCAATTTCAAAATTTA 1693  
Qy 197 ThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGly 216  
Db 1694 GGTAATATTCATGCTAGTATAATAATCTTAAATAGATGGTTGTAGATACA--- 1750  
Qy 217 LeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAla 236  
Db 1751 -----CATAGATATATTTGGATAAAATATTTTAAATCTTTTGTAT 1789  
Qy 237 LysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256  
Db 1790 AAGGAATTAATGAAAGAAATCAAGATTTATATGATTAATCAATCAATCAATCGGTAT 1849  
Qy 257 ValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetValAsnIle 276  
Db 1850 TAAAGACTTTTGGGTGATTATTACATATGATAAACCACTATATGTTAAATTTA 1909  
Qy 277 AspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
Db 1910 TATGATCAAAATAATATGCGATGGAATAATGATAGGTATATAGAGGTATATGATCATT 1969  
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnGlu 303  
Db 1970 AAAGGCGCTAGAGTAGGTAATCAGTACACAAACATTTATTTAAATCAAGTTGTATAGG 2029  
Qy 304 GlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGly 323  
Db 2030 GGCAGAAATTTATATAAAATAATGCTCTCGAAATAAGATATAATTTGTAGAAAT 2089  
Qy 324 GlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLys 343  
Db 2090 ATGATCGTGTATATTAATGATAGTAGTAAATAAATAAAGAAATATAGGTTA----- 2140  
Qy 344 AsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArg 363  
Db 2141 -----GCTACPAATGCATCAGCAGCGCGTAGAATAAATACTAAGTGA 2185  
Qy 364 GluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThr 383  
Db 2186 TTAGAAATACCTGATGAGGAAATCTAAGTCACTAGTACTAGTAAAGTCAAAAATGAT 2245  
Qy 384 TyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIle 403  
Db 2246 CAAGGAATAACAAATAAATGTC---AAATGAATTTACAAGATAATAATGGGAATGATATA 2302  
Qy 404 CysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyrLeu 423  
Db 2303 GCCTTTATAGGATTCATCAGTTT-----AATAATATA 2335  
Qy 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
Db 2336 GCTAAAGTAGTAGCAAGTAATTTGGTATAATAGACAAATAGAAAGATCTAGTAGGCTTTG 2395  
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450

Db 2396 GGTGCTCATGGGAATTTATCTCTGTAGATGAT 2428  
RESULT 12  
US-10-051-952-3  
; Sequence 3, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 09/730,237  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: botulinum toxin  
US-10-051-952-3  
Alignment Scores:  
Pred. No.: 9,99e-54 Length: 3876  
Score: 561.00 Matches: 159  
Percent Similarity: 50.21% Conservative: 84  
Best Local Similarity: 32.85% Mismatches: 173  
Query Match: 23.46% Indels: 68  
Gaps: 21  
US-09-910-186a-10 (1-450) x US-10-051-952-3 (1-3876)  
Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsnGlu 22  
Db 2506 ATGCGTTTGCATCTTCAATATATACCAATGATACAACTACTAATAAGAAATGTTAATAAA 2565  
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThrLeu 42  
Db 2566 TATAATAGCGAAATTTTAAATAATATATCTTAAATTAAGATATAAGATATAATTTA 2625  
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
Db 2626 ATAGATTTATCAGGATATGGGCAAGGTAGAGGTATATGATGAGTCTGAGCTTAATGAT 2685  
Qy 63 IlePheProPheAspPheLysLeuGlySerSerGlyLysArgGlyLysValIleVal 82  
Db 2686 -----AAAAATCAATTTAAATTAAGTACTAGTTCAGCA-----AATAGTAGATAGAGTG 2733  
Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102  
Db 2734 ACTCAAAATCAGATATACATATTTAATAGTGTCTCTTCTGATTAGCGTTAGCTTTGG 2793  
Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116  
Db 2794 ATAAGAATACCTAAATATAAGAATGATGGTATACAAAATATATATTCATAATGAATATACA 2853  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 2854 ATAAATTAATCTGTAAGAAATAATATTCGGCTGGAATAATCTATTAGGGTAAATAGATA 2913  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
Db 2914 ATATGGACTTTAATTGATATAAATGGAATAAATGGAATAAATGGAATAAATGGAATAA 2973  
Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
Db 2974 AGAAGATATATCAGAGTATATAAATAGATAGTGTCTTCTGTAATACTAATAAT--- 3030  
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuLeuAspThrIleLysValLysGlu 195  
Db 3031 TTGAATAACGCTAAATTTATATTAATGTTAGTAAAGTAGAATCAAAATCAAGATATAAAGAT 3090





```

Qy 249 AsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268
Db 699 GACACCAAGTCCCAATTCGTGTCTGTAAGAGAGCTTGGGGTGACCTACCTGAGTACGAC 758
Qy 269 LysGluTyrTrpMetValAsnLeuAspTyrLeuAsnArgTyrMetTyrAlaAsnSer--- 287
Db 759 AACCGTACTACATCGTGAATCTGACATCCGACCAAAATACGTTGACGTCACCAATGTA 818
Qy 288 -----ArgGlnIleValPheAsnThrArg--- 295
Db 819 GGTATCCGCGGTTTACATGTACCTGAAAGGTCGCGGTCTCTGTATGACTACCAACATC 878
Qy 296 ArgAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsn 315
Db 879 TACCTGACTCTCCCTGCTACCGGTGACCAATTCATCATCAAGAAATACCGCTGCTGT 938
Qy 316 ThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsn 335
Db 939 AACAGGACAATATCGTTCGCAACAATGATCGGTATACATCAATGTTGTAGTTAAGAAC 998
Qy 336 LysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355
Db 999 AAGAATACCGCTCG-----GCTACCAATGCTTCTCAGGCT 1034
Qy 356 AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375
Db 1035 GGTGTAGAAAAGATCTGCTGCTGCTGGAATCCCGACGTTGGTAACTGCTCAGGTA 1094
Qy 376 GlnIleGlnPrometAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe 395
Db 1095 GTTGTAATGAATCCAAAGACGACCGGATCATCAACAAATGC---AAATGAATCG 1151
Qy 396 AsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGly 415
Db 1152 CAGGACACATATCGTACGATATCGGTTTCATCGGTTCCACCATC----- 1199
Qy 416 AspTyrPyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435
Db 1200 -----AACAAATGCTGAATCGTGGTCTCTGGAGTCAATCCCGTGTGATGAC 1301
Qy 436 LeuGluSerThrSer-----ThrHisTrpGlyPheValProValSerGlu 450
Db 1245 ATCGAAGCTTCTCTCGCACTCTGGTGTCTCTGGAGTCAATCCCGTGTGATGAC 1301
RESULT 14
US-10-051-952-1
; Sequence 1, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3891
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-1
Alignment Scores:
Pred. No.: 7,13e-53 Length: 3891
Score: 553.50 Matches: 146
Percent Similarity: 48.31% Conservative: 82
Best Local Similarity: 30.93% Mismatches: 189
Query Match: 23.15% Indels: 55
DB: 12 Gaps: 11
US-09-910-186a-10 (1-450) x US-10-051-952-1 (1-3891)

```

```

Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
Db 2545 ATACCTTTTCAGGTTTCCAAATAGGTAGATAATCAAGATTATTAATCTCATTTACTGAA 2604
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
Db 2605 TATATTAAAGAAATATTATTAATCTCTATATTGAATTAAGATATGAAAGTAATCATTTA 2664
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
Db 2665 ATAGACTTATAGGTATGATCAAAAATAAATATTGCTAGTAAAGTAATTTTGTACCA 2724
Qy 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
Db 2725 ATAGATAAAATCAATTCATTAATTTAGAAAGTAGT-----AAA 2769
Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99
Db 2770 ATTGAGGTAATTTTAAAAAATGCTATTGTATATAAATAGTATGATGAAAAATTTTACT 2829
Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116
Db 2830 AGCTTTTGGATAAGAAATCTCTAAGTATTTTACAGATTAAGTCTAAATTAATGAATATACA 2889
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 2890 ATAATAATTTGATGGAATAATAATTCAGGATGGAAGATGATCACTTAATTTGTTGATAA 2949
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 2950 ATCTGGGACTTTACAGGATCTCAGGAAATAAACAAGAGTAGTATTTTAAATAGAGTCAA 3009
Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 3010 ATGATTATATATACAGATTATATAACACAGATGATTTTGTAACTACTCACTAATAATA 3069
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysIleAspThrIleLysValLysGlu 195
Db 3070 TTAATAACTTAAATTTATATAATGGAAGATTATAGATGAAACCAATTTCAAT 3129
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3130 TTAGTAAATTCATGCTAGTAAATATAATATGTTTAAATAGATGTTGTAGAGATACA 3189
Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
Db 3190 -----CATAGATATATTTGGATAAAATATTTTAACTCTTTT 3225
Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255
Db 3226 CATAGGAATTAATGAAGAAAGAAATCAAGATTATATGATCAATCAATTCAGGT 3285
Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 3286 ATTTTAAAGACTTTTGGGTGATTTATTTACAATATGATAAACCATAGTATATGTTAAAT 3345
Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 3346 TTATATGATCCAAATAATATGATGATGATGATGATGATGATGATGATGATGATGAT 3405
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
Db 3406 CTTAAAGGCGCTAGAGGTAGGTAATGACTACAAACATTTTATTAATTCAGTTGTAT 3465
Qy 303 GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
Db 3466 AGGGGACAAAATTTTATATAAAAAATATGCTTCGGAATTAAGATATATATTGTTAGA 3525
Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
Db 3526 AATAATGATCGTGTATATATTAATGTTAGTAGTTTAAATAAAGAAATATAGGTTA----- 3579

```



```

QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
Db 3560 -----GCTACTAATGATCAGAGGCGGTAGAAAAAATACTAAGT 3621
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
Db 3622 GCATTAGAAATACCTGATAGAAATCTAAGTCAAGTAGTAGTAATCAAGTCACAAAAAT 3681
QY 383 ThrTyrTyrThrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 3682 GATCAAGGAATACAAATAAATGC---AAATGAATTTACAGATAAATAATGGGAATGAT 3738
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyr 422
Db 3739 ATAGGCTTTATAGGAATTCATCAGTTT-----AATAAT 3771
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 3772 ATAGCTAAAGTAGTCAAGTAATGGTATAATAGACAAATAGAAAGATCTAGTAGGACT 3831
QY 441 -----ThrHisTyrGlyPheValProValSerGlu 450
Db 3832 TTGGGTGCTCATGGGAATTTATCTCTGTAGATGAT 3867

RESULT 15
US-09-350-756-9
; Sequence 9, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; EARLIER FILING DATE: 1999-07-09
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 9
; LENGTH: 1327
; TYPE: DNA
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-9

Alignment Scores:
Pred. No.: 3,75e-53 Length: 1327
Score: 550.00 Matches: 146
Percent Similarity: 48.39% Conservative: 80
Best Local Similarity: 31.26% Mismatches: 174
Query Match: 23.00% Indels: 67
DB: 10 Gaps: 12

US-09-910-186a-10 (1-450) x US-09-350-756-9 (1-1327)
QY 16 LeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeu 35
Db 10 TTATATCTCATCTTACTGATATATAGAAATATTATTAATCTCTTCTATATGAATTTA 69
QY 36 GlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 55
Db 70 AGATATGAAGTATCATCTTTATAGACTTATCTAGGTATGCATCAAAATAAATATTGGT 129
QY 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72
Db 130 AGTAAAGTAATTTTGTATCCAAATAGATAAAAAATCAAATTCATTTATTTAATAGAAAGT 189

```

```

QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer 92
Db 190 AGT-----AAAATTCAGGTAATTTTAAAAATGCTATTTGTATATAATAGT 234
QY 93 MetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValSerAsnLeu 112
Db 235 ATGTATGAAATTTTGTACTAGCTTTTGGATAGAAATCTCTAAGTATTTTAAACAGTATA 294
QY 113 Pro-----GlyTyrThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129
Db 295 AGTCTAAATTAATGAATATACATAATATAATGTATGGAAAAATAATTCAGGATGGAAGTA 354
QY 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149
Db 355 TCACCTTAATTTATGGTGAATAATCTGGACITTTACAGGACTCTCAGGAATAAACAAGA 414
QY 150 IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr---AsnLysTrpPhePhe 168
Db 415 GTAGTTTATAATACAGTCAATGATTAATATATATCATGATTAATAACACAGATGATTTT 474
QY 169 ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle 188
Db 475 GTAACATATCATAATAATAGATTAATAAATCTAAATTTATATAAATGGAAGATTATA 534
QY 189 AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu 208
Db 535 GATCAAAACCAATTTCAATTTAGTAAATATTCAGTAGTAATAATAATGTTTAAA 594
QY 209 IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp 228
Db 595 TTAGATGGTGTAGAGATACA-----CATAGATATATTTGG 630
QY 229 IleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe 248
Db 631 ATAAATATTTAATCTTTTGTATAGGAATTAATGAAAAAGAAATCAAGATTTATAT 690
QY 249 AsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268
Db 691 GATAATCAATCAATTCAGGTATTTTAAAAGACTTTTGGGTGATTTTACAATATGAT 750
QY 269 LysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer--- 287
Db 751 AAACCATATATATATGTTAAATTTATATGATCCAAATAATATGTCGATGTAATAATGTA 810
QY 288 -----ArgGlnIleValPheAsnThrArg 296
Db 811 GGTATTAGAGTTATATGATCTTAAAGGCGCTAGAGTAGCGTAACTACTACA----- 864
QY 297 AsnAsnAsnAspPhe-Asn-----GluGlyTyrLysIleI 308
Db 865 -----AATGACTACAAACATTTTAAATTCAGTTTGTATAGGGGACAAATTTAT 918
QY 308 eileLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTy 328
Db 919 TATAAAAAATATGCTTCGGAATAAAGATAATATTTGTTAGAAATAATGATCGTGATA 978
QY 328 rPheAspMetThrIleAsnAsnLysAlaIyrAsnLeuPheMetLysAsnGluThrMetTy 348
Db 979 TATTAATGTAGTAGTTTAAATAAAGAAATATAGGTTA----- 1015
QY 348 rAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAs 368
Db 1016 -GCTACTATGCTACACAGCGCGCTAGAAAAAATACTAAAGTCATAGAAATACCTGA 1074
QY 368 pIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSe 388
Db 1075 TGTAGGAAATCTAAGTCAAGTAGTAGTAATGAAGTCAAAAAATGATCAAGGAATAACAAA 1134
QY 388 rGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyTh 408
Db 1135 TAAATGC---AAATGAATTTACAAATTAATATGGAATGATATAGGCTTTATAGGATT 1191
QY 408 rTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyrLeuValProThrVally 428

```

```

Db 1192 TCATCAGTTT-----|||   |||
QY 428 scInGlyAsnTyAlaSerLeuLeuGluSerThrSer-----ATAATATAGCTAAACTAGTAGC 1224
Db 1225 AAGTAATTGGTATATAGACAAATAGAAAGATCTAGTAGGACTTTGGGTTGCTCATGGGA 1284
QY 444 yPheValProValSerGlu 450
Db 1285 ATTTATCTCTGTAGATGAT 1303

```

Search completed: November 7, 2002, 20:06:56  
 Job time : 86 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 17:48:22 ; Search time 3151 Seconds  
(without alignments)  
3590.644 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPNFSTNNLSLLKDII.....NYASLLESTHWGFPVNSE 450

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002.111618\_20892/app\_query.fasta\_1.647  
-DB=Pending\_Patents\_NA\_Main -QFMT=fastap -SUFFIX=rnrm -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MNLLEN=0 -MAXLEN=2000000000  
-USER=US09910186.ecgn.1.1.1396 -runat\_04112002.111618\_20892 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_XMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

Pending\_Patents\_NA\_Main: \*  
1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq: \*  
7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq: \*  
8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq: \*  
9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq: \*  
10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq: \*  
11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq: \*  
12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq: \*  
13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq: \*  
14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq: \*  
15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq: \*  
16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq: \*  
17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq: \*  
18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq: \*  
19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq: \*  
20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq: \*  
21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq: \*  
22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq: \*  
23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq: \*  
24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq: \*  
25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq: \*  
26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq: \*  
27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq: \*  
28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq: \*

29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq: \*  
30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq: \*  
31: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq: \*  
32: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq: \*  
33: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq: \*  
34: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq: \*  
35: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq: \*  
36: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq: \*  
37: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq: \*  
38: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq: \*  
39: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq: \*  
40: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq: \*  
41: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq: \*  
42: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq: \*  
43: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq: \*  
44: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq: \*  
45: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq: \*  
46: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq: \*  
47: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq: \*  
48: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq: \*  
49: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq: \*  
50: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq: \*  
51: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq: \*  
52: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq: \*  
53: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq: \*  
54: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq: \*  
55: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq: \*  
56: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq: \*  
57: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq: \*  
58: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq: \*  
59: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq: \*  
60: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq: \*  
61: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq: \*  
62: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq: \*  
63: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq: \*  
64: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq: \*  
65: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq: \*  
66: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq: \*  
67: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq: \*  
68: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq: \*  
69: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq: \*  
70: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq: \*  
71: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq: \*  
72: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq: \*  
73: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq: \*  
74: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq: \*  
75: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq: \*  
76: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq: \*  
77: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq: \*  
78: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq: \*  
79: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq: \*  
80: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq: \*  
81: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq: \*  
82: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq: \*  
83: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq: \*  
84: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq: \*  
85: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2391	100.0	1371	23	US-09-611-419A-9
2	2391	100.0	1371	34	US-09-910-186A-9
3	2386	99.8	3876	1	PCT-US97-15394-59
4	2386	99.8	3876	11	US-08-704-159-59
5	2386	99.8	3876	42	US-10-205-516-5
6	2386	99.8	3876	43	US-10-271-012-59

```
7 2386 99.8 3906 42 US-10-205-516-19
8 2386 99.8 3950 13 US-08-954-202-2
9 2381 99.6 1371 17 US-09-350-756-3
10 2364 98.9 3876 29 US-09-730-237-4
11 2364 98.9 3876 39 US-10-051-952-4
12 2319 97.0 1502 1 PCT-US97-15394-61
13 2319 97.0 1502 11 US-08-704-159-61
14 2319 97.0 1502 43 US-10-271-012-61
15 868.5 36.3 1374 23 US-09-611-419A-11
16 867.5 36.3 3831 1 PCT-US97-15394-65
17 867.5 36.3 3831 11 US-08-704-159-65
18 867.5 36.3 3831 42 US-10-205-516-7
19 867.5 36.3 3861 43 US-10-271-012-65
20 867.5 36.3 3861 42 US-10-205-516-21
21 865.5 36.2 3952 21 US-09-547-188-7
22 865.5 36.2 1374 34 US-09-910-186A-11
23 863.5 36.1 3831 29 US-08-730-237-5
24 863.5 36.1 3831 39 PCT-US97-15394-5
25 803.5 33.6 1469 1 PCT-US97-15394-67
26 803.5 33.6 1469 11 US-08-704-159-67
27 803.5 33.6 1469 43 US-10-271-012-67
28 604.5 25.3 3894 1 PCT-US97-15394-76
29 604.5 25.3 3894 11 US-08-704-159-76
30 604.5 25.3 3894 42 US-10-205-516-13
31 604.5 25.3 3894 43 US-10-271-012-76
32 604.5 25.3 3924 42 US-10-205-516-27
33 604.5 25.3 3937 21 US-09-547-188-13
34 602.5 25.2 3759 42 US-10-205-516-9
35 602.5 25.2 3789 42 US-10-205-516-23
36 595.5 24.9 3753 1 PCT-US97-15394-49
37 595.5 24.9 3753 11 US-08-704-159-49
38 595.5 24.9 3753 43 US-10-271-012-49
39 594.5 24.9 1400 34 US-09-910-186A-13
40 594.5 24.9 3759 1 PCT-US97-15394-51
41 594.5 24.9 3759 11 US-08-704-159-51
42 594.5 24.9 3759 43 US-10-271-012-51
43 594.5 24.9 4030 21 US-09-547-188-5
44 594.5 24.9 4030 21 US-09-547-188-9
45 593 24.8 3876 1 PCT-US97-15394-39
```

## ALIGNMENTS

## RESULT 1

```
US-09-611-419A-9
; SEQUENCE 9, Application US/09611419A
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard A.
; APPLICANT: Byrne, Michael P.
; APPLICANT: Middlebrook, John L.
; APPLICANT: Lapenotiere, Hugh
; APPLICANT: Clayton, Michael A.
; APPLICANT: Brown, Douglas R.
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; FILE REFERENCE: A3626 067252.0105
; CURRENT APPLICATION NUMBER: US/09/611.419A
; CURRENT FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/146,192
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 42
```

```
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on BONTA Hc
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-611-419A-9
Alignment Scores:
Pred. No.: 4,07e-230 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-910-186A-10 (1-450) x US-09-611-419A-9 (1-1371)
QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20
Db 10 ARGACATCCCATTCACATCTTCTCTACACCAACCACTCCCTGTTGAAGGACATCATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTAGTCTTCAACACATCAACGACTCCAGATCCCTGCTCCCTGCAGAACCCGTAAGA 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTCGACACCTCCGGTTACACCCCGAGGTCTCCGAGGAGGGTGCAGTCCAGCTG 189
QY 61 AsnProIlePheProPheAspPheLysLeuGlySerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCTCCGCTGAGGACAGAGTAAGGTC 249
QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Db 250 ATCTGTCACCCAGACGACGACATCTGCTACACTCCATGACGAGTCTCTTCCATCTCC 309
QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
Db 310 TTCTGGATCAGAATCAACAAGTGGTCTCCAACTTGCAGGTTCACCATCATCATCGACTCC 369
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGAACAACTCCGGTGGTCCATCGGTATCATCTCCAACTTCTCTGGTCTTCCACCTG 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 430 AAGCAGACGAGGAGTCCGAGCAGTCCATCAACTTCTCTACGACATCTCCACACAGCT 489
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CTGTGTTACAAAGTGGTCTTCTGTCACCTCCACCAACAATGTTGGGTAACTAAGAG 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Db 550 ATCTACATCAACGGTAAGTGTATCGACACCATCAAGGTCAAGGAGTTCACCGGTATCAAC 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCAGACCATCACCTTCGAGATCAACAGATCCCAAGATCCCGACACCGGTCTGATCACCTCC 669
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheThrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCCGACACATCAACATGTGGATCCGCTGACTTCTACATCTTCTCGCAAGGAGTTCGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnValValLysAspTyr 260
Db 730 GGTAAGGACATCAACATCTTCTCAACTCTTTCAGTACACCAACGCTCGTCAAGGACTAC 789
```

```
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
|
|
|
Db 790 TGGGGTAAGACCTGAGATACAAAGAGTACTACATGGTCAACATCGACTACTTGAAC 849
|
|
|
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsp 300
|
|
|
Db 850 AGATACATGTAGCGCAATCCAGACAGATCTCTCTACACACGACGTAAACAACGAC 909
|
|
|
QY 301 PheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
|
|
|
Db 910 TTCACGAGGGTTACAGATCATCATCAAGGTATCAGAGGTACACACACGACACAGA 969
|
|
|
QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
|
|
|
Db 970 GTCAGAGGTGGTACATCTGTCTACATGACATGACTATCAACAACAAGGCGCTCAACCTG 1029
|
|
|
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
|
|
|
Db 1030 TTCATGAGAAGACGACCATGATCCGCGACACACCATCCACCGAGGACATCTAGCGCATC 1089
|
|
|
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
|
|
|
Db 1090 GGTCTGCGTGAACAGACATCAACAGACATCAACACATCATCTCCAGATCCAGCCAAATG 1149
|
|
|
QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
|
|
|
Db 1150 AACACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATC 1209
|
|
|
QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHis 420
|
|
|
Db 1210 TCCGGTATCTGTTCCATCGGTACCTACAGATTCGGTCTGGTGGTGACTGTGACAGACAC 1269
|
|
|
QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
|
|
|
Db 1270 AACTACTTGGTTCCTCAACTGTCRAGCAGGTAACACTACGCTCTCTGCTGGAGTCCACITCC 1329
|
|
|
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
|
|
|
Db 1330 ACCCACTGGGGATTGCTCCAGTCTCCGAG 1359
|
|
|
RESULT 2
US-09-910-186A-9
; Sequence 9, Application US/09910186A
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Materiel Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; FILE REFERENCE: A33626-A 067252, 0107
; CURRENT APPLICATION NUMBER: US/09/910,186A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-05
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; OTHER INFORMATION: sequence
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-910-186A-9
```

```
Alignment Scores:
Pred. No.: 4,07e-230 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 34 Gaps: 0
```

US-09-910-186A-10 (1-450) x US-09-910-186A-9 (1-1371)

```
QY 1 MetThrIlePhePheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20
|
|
|
Db 10 ATGACATCCCATTCACATCTTTCCTTACACCAACAACTCCCTGTTGAAGGACATCATC 69
|
|
|
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
|
|
|
Db 70 AACGAGTACTTCAACAACATCAACGACTCCAAAGATCCTGTCCCTGCAGAACCGTAAGAAC 129
|
|
|
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
|
|
|
Db 130 ACCTTGGTGGACACCTCCGGTTACAACCCGAGGTCTCCGAGGAGGTGACGCCAGCTG 189
|
|
|
QY 61 AsnProIlePhePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
|
|
|
Db 190 AACCCAACTTCTCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGAGTAAAGTCC 249
|
|
|
QY 81 IleValThrGlnAsnGluAsnIleValIleValIleValIleValIleValIleValIle 100
|
|
|
Db 250 ATCGCACCCAGACAGACATCGTCTACAACTCCATGTACGAGTCTCTTCCATCTCC 309
|
|
|
QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
|
|
|
Db 310 TTCGTGATCAGATCAACAAGTGGTCTCCAACTGCCAGGTACACCATCATCGACTCC 369
|
|
|
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
|
|
|
Db 370 GTCAAGACACACTCCGGTGGTCCATCGGTATCATCTCCAACCTCTTGGTCTTCCACCTG 429
|
|
|
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
|
|
|
Db 430 AAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCCAAACAGCT 489
|
|
|
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
|
|
|
Db 490 CCTGGTTACAACAAGTGGTCTTCGTCACCGTCAACAACAACATGATGGTAAACATGAAG 549
|
|
|
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
|
|
|
Db 550 ATCTACATCAACGGTAAAGTGTATCGACACCATCAAGGTCAAGGAGTTCACCGGTATCAAC 609
|
|
|
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
|
|
|
Db 610 TTCTCCAAGACCATCACCTTCGAGATCAACAAGATCCCAAGATCCACCGGTCTGATCACCTCC 669
|
|
|
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240
|
|
|
Db 670 GACTCCGACAAACATCAACATGTGGATCCGTGACTTCTACATCTTCGCCAAGAGGATGGAC 729
|
|
|
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260
|
|
|
Db 730 GGTAAAGGACATCAACATCCTGTTCAACTCCTTGCAGTACACCAACGCTCGTCAAGGACTAC 789
|
|
|
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
|
|
|
```

```
Db 790 TGGGTAACGACCTGAGATACAAAGGAGTACTACATGCTCAACATCGACTACTTGAAC 849
Qy 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTACGCACTCCAGACAGATCGTCTTCAACACACAGAGTAAACACACGAC 909
Qy 301 PheAsnGluGlyTyrLysIleIleLysArgGlyAsnThrAsnAspThrArg 320
Db 910 TTCACGAGGGGTTACAAGATCATCATCAAGCGTATCAGAGGTAAACACACACACGAC 969
Qy 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Db 970 GTCAGAGGTGTGACATCCCTGCTACTTCGACATGACTATCAACACACAGCCCTACACCTG 1029
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1030 TTCATGAAGACGAGACCATGTACGCGACACCACTCCACCGGAGGACATCTACGCCATC 1089
Qy 361 GlyLeuArgGluGlnThrLysAspIleAspAsnIleIlePheGlnIleGlnProMet 380
Db 1090 GGTCTGGTGACGACAGACAGGACATCAACACACACATCATCTCCAGATCCAGCCCATG 1149
Qy 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1150 AACACACATCTACTACTACGCTTCCAGATCTTCAAGTCCAACTCAACGGTGAGAACATC 1209
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyClyAspThrArgHis 420
Db 1210 TCCGGTACTCTTCCATCGGTACTACAGATCCGCTGGTGGTGACATGGTACAGACAC 1269
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1270 AACTACTGTGTCCAACTGTCAACGAGGGTAACAGCCCTCCCTGCTGGAGTCCACTTCC 1329
Qy 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 1330 ACCACTGGGATTCGCTCCAGTCCGAG 1359

RESULT 3
PCT-US97-15394-59
; Sequence 59, Application PC/TUS9715394
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/15394
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs.
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3873
; PCT-US97-15394-59

Alignment Scores:
Pred. No.: 4,77e-229 Length: 3876
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 1 Gaps: 0

US-09-910-186a-10 (1-450) x PCT-US97-15394-59 (1-3876)
Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21
Db 2327 ACAATACCCCTTAAATATTTTTCATATACTAATAATCTTTATAAAGATATAATAT 2585
Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2587 GAATATTTCAATATATATGATTCAAAATTTTTCAGCCTACAAAACAGAAAAATACT 2646
Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2647 TTAGTGATACATCAGGATATAATGCAAGTGTAGTGAAGAGGCGATGTTCCAGCTTAT 2706
Qy 62 ProfilePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGTGTAGTTTCAGGGGAGGATAGAGTTAACTATA 2766
Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 2767 GTAACCCAGAAATATATGATATAATTTCTATATGATGAAAGTTTATGCAATAGTTT 2826
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2827 TGGATTAGATAATAATRAATGGTAAAGTAAATTTTACCCTGGATATATAATATGATGTT 2886
Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 2887 AAAATAACCTCAGGTTGGAGTATAGTATATTAGTAAATTTTATGATTTTACTTTAAA 2946
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161
Db 2947 CAAATCAAGATAGTAGAACAAAGTAAATTTTAGTATGATATATCAATATGCTCT 3006
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3007 GGATACAAATAATGGTTTTTTTGTAACTGTACTAAACAATATGATGGAAATATGAAGAT 3066
Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3067 TATATAATGAAATTAATTAAGATATATAAAGTTAAAGAACTAACTGGAATTAATTT 3126
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3127 AGCAAACTATACATTTGAATAATAATAAAATCCAGATACCGGTTTGATTTACTTCAGAT 3186
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3187 TCTGATACATCAATATGTGGTAAGAGATTTTATATATTTCCTAAAGAAATAGATGT 3246
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261
Db 3247 AAGATATATATATATTTAATAGCTTCCATATACATAATGTTGCTAAAGAAATATTGG 3306
Qy 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281
```

```
Db 3307 GCAATGATTTAAGATATATAAAGATATATATATGTTAATATATAGTATTTAATAATAGA 3366
Qy 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301
Db 3367 TATATGATGGAACCTCAGCAGCAATTTGTTTTAATACACGTAGAAATATATATGACTTC 3426
Qy 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspPheArgVal 321
Db 3427 AATGAAGATATAAATATATAAATAAAGAAATCAGAGAAATACAAATGATAGTAGTA 3486
Qy 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3487 CGAGGAGAGATATTTATATTTGATGACATTAATACAAACGATATATTTGTTT 3546
Qy 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3547 ATGAAGATGAACATATGATGACAGATTAATCAATCAAAATTTTAAATGAGATATATGCTATAGT 3606
Qy 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3607 TTAGAGAACAAACAGGATATAATATGATATATATATATTTCAATACACCAATGAAT 3666
Qy 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 3667 ATACTTATATACGATCTCAATATATTAATCAAAATTTTAAATGAGAAATATTTCT 3726
Qy 402 GlyIleCysSerIleGlyThrTyrArgPheArgGlyGlyAspTyrArgHisAsn 421
Db 3727 GGAATATGTTCAATAGGACTATTCGTTTATAGACTTGAGAGTATGATATAGACAAAT 3786
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThr 441
Db 3787 TATTTGGTGCCTACTGTGAGCAAGAGAAATATGCTTCATTATATAGATCAACATCAACT 3846
Qy 442 HisTrpGlyPheValProValSerGlu 450
Db 3847 CATTTGGGTTTGTACCTGTAAAGTAA 3873

RESULT 4
US-08-704-159-59
; Sequence 59, Application US/08704159
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08704,159
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3873
US-08-704-159-59

Alignment Scores:
Pred. No.: 4,77e-229 Length: 3876
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 11 Gaps: 0

US-09-910-186A-10 (1-450) x US-08-704-159-59 (1-3876)
Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2527 ACAATACCCCTTAAATATTTTTCATATACATAATAATCTTTATTAAGATATAATAT 2586
Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2587 GAATATTTCAATAATATTAATGATTCAAAAATTTTGAGCCTACAAACACAGAAAAATACT 2646
Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2647 TTATGATACATCAGGATATATATGCAGAAAGTGTGAAGAGCGGATGTTGACCTTAAT 2706
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGGTAGTTTCAGGGGAGGATAGAGTAAAGTTATA 2766
Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrClnSerPheSerIleSerPhe 101
Db 2767 GTACCCAGAAATGAAATATTTGATATATATTTCTATGATGAAAGTTTTCAGCATAGTTT 2826
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2827 TGGATTAGAATAAATAAATGGTAAAGTAAATTTTACCTGGATATATCTATAATGATAGTT 2886
Qy 122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 2887 AAAATAACACAGGTTGGAGTATAGGTATTATTAGTAATTTTATTAGTATTACTTTTAAA 2946
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
Db 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAATAATGCTCCT 3006
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3007 GGATACAAATGAATGTTTTTTTGTAACTGTTTACTAAACAATATGATGGAAATATGAAGAT 3066
Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3067 TATATAATGGAATTAATAGATACATATAAAGTTAAAGAACTAACTGGAATTAATTTT 3126
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3127 AGCAAAACATATAACATTTGAATAAATAAATTTCCAGATACCGGTTTGATTTCTCAGAT 3186
Qy 222 SerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3187 TCTGATACATCAATATGTGGATAAGAGATTTTATATATTCTTAAGAAATAGATGCT 3246
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261
Db 3247 AAAGATATTAATATATTAATTAATAGCTTCAATATACTAATGTTGTAAGAGATTTATGG 3306
Qy 262 GlyAsnAspLeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrIleuAsnArg 281
```







```
Db 3427 ATGAAGGATATAAAATATATAATAAAGAAATCAGAGGAATACAAATGATCTAGAGTA 3486
QY 322 ArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3487 CGAGGAGGAGATATTTATATTTTATGATGACAAATTAATAACAAGCATATAATTTGTTT 3546
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3547 ATGAAGAATGAACACTATGTCACATANTCATAGTACTGAAGATATATATGCTATAGGT 3606
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3607 TTAAGAGAAACAACAAAGATATAAATGATATATATATATTTCAATACACCAATGAAT 3666
QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 3667 ANACTTATATATAGCATCTCAATATATTTAAATCAATTTTATGAGAGAAATATTTCT 3726
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421
Db 3727 GGAATATGTTCAATAGTACTTATCGTTTATAGACTTGGAGGTGATGCTATAGACACAAT 3786
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3787 TATTTGGTCCCTACTGTGAAGCAAGAAATATGCTTCATTTATGAAATCAACATCAACT 3846
QY 442 HisTyrGlyPheValProValSerGlu 450
Db 3847 CATTTGGGGTTTGTACCTGTAAGTGA 3873
```

## RESULT 7

```
US-10-205-516-19
; Sequence 19, Application US/10205516
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; FILE REFERENCE: J2btxl
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 3906
; TYPE: DNA
; ORGANISM: Clostridium botulinum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3906)
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: (1347)..(1358)
; OTHER INFORMATION: factor xa site
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: (3886)..(3903)
; OTHER INFORMATION: 6-histidine tag
US-10-205-516-19
```

```
Alignment Scores:
Pred. No.: 4, 82e-229 Length: 3906
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 42 Gaps: 0
```

US-09-910-186a-10 (1-450) x US-10-205-516-19 (1-3906)

```
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2539 ACAATACCCCTTTAATATTTTTCATATACATAATATCTTCTTATTTAAAGATATAATTAAT 2598
```

```
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2599 GAATATTTCAATATATATTAATGATTTCAAAATTTTGACCTACAAAACACAGAAAATAACT 2658
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2659 TTAGTGGGATACATCAGGATATAAATGCAGAGTGAAGTGAAGGCGATGTTCAGCTTAAT 2718
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2719 CCAATATTTCCATTTGACTTTAAATTTAGTGTTCAGGGAGGATAGAGTAAAGTTATA 2778
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 2779 GTAACCCAGAAATGAATATGTTATATATCTATGATGAAGCTTTTAGCATTTAGTTT 2838
QY 102 TyrIleArgIleAsnLysTyrPheValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2839 TCGATTAGAATAAATAAATGGGTAAATTTAGTGTGATATATCTAATAATGATAGTGT 2898
QY 122 LysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 2899 AAAATAACTCAGGTTCAGATATAGTATATTTAGTAATTTTGTAGTATTTACTTTAAA 2958
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161
Db 2959 CAAAATGAAGATAGTAGCAAAAGATATAAATTTTATGTTATGATATATCAATAATGCTCT 3018
QY 162 GlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3019 GGATACAAATAATGGTTTTTGTAACTCTTACTAACATATGATGGCAATATGAAGATT 3078
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3079 TATATAAATGGAAATTAATAGATACTATAAAGCTTAAAGAACTAACTGGAAATAATTT 3138
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProaspThrGlyLeuIleThrSerAsp 221
Db 3139 AGCAAACTATPAAACATTTGAATAAATAATTCAGATACCGGTGTGATCTACTTCAGAT 3198
QY 222 SerAspAsnIleAsnMetTyrPheIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3199 TCTGATAACATCAATATGTGGATAGAGATTTTATATATTTGCTAAAGAAATTAGATGGT 3258
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTyr 261
Db 3259 AAAGATATTAATATATATTTAAATAGCTTGCATATATCTAATATGTTGTTAAAGATTATGG 3318
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281
Db 3319 GGAAATGATTTAAGATATAAATAAAGAAATATATATGTTTAAATATAGATTTATTAATAATAGA 3378
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301
Db 3379 TATATGATGCCAACTCCGACAAATTTGTTTAAATACACGTAAGAAATAATATGACTTC 3438
QY 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 3439 AATGAAGGATATAAATAATATAAATAAAGAAATCAGAGAAATACAAATGATCTAGAGTA 3498
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3499 CGAGGAGGAGATATTTTATTTTATGATGACAAATTAATAACAAGCATATAATTTGTTT 3558
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3559 ATGAAGAATGAACACTATGTTATGCAATATCATAGTACTGAAGATATATATGCTATAGGT 3618
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3619 TTAAGAGAAACAACAAAGATATAAATGATAATATTTATTTCAATACCAATGAAT 3678
```





```
US-09-730-237-4
; Sequence 4, Application US/09730237
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Treating Hyperhidrosis
; FILE REFERENCE: 2933
; CURRENT APPLICATION NUMBER: US/09730,237
; CURRENT FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-09-730-237-4
Alignment Scores:
Pred. No.: 7,83e-227 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservatives: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 29 Gaps: 0
US-09-910-186a-10 (1-450) x US-09-730-237-4 (1-3876)
QY 2 ThrileProPheAsnIlePheSerTyrrAsnSerLeuLeuLysAspIleIleAsn 21
DB 2527 ACAATACCTTTAAATATTTTCATATACTAATAATCTTTATTAAGAATATAATTAAT 2586
QY 22 GluTyrrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
DB 2587 GAATATTTCCAAATATTAATGATTCAAAATTTTGAGCCCTACAAAACAGAAAATACT 2646
QY 42 LeuValAspThrSerGlyTyrrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
DB 2647 TTAGTGGATACATCAGGATATAATGACAGAGTGTGAGAGAGCGATGTCAGCTTAAT 2706
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
DB 2707 CCAATATTTCCATTTGACATTTAAATAGTAGTACGAGGAGGAGGAGGAGGAGGAGGAGG 2766
QY 82 ValThrGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPhe 101
DB 2767 GTAACCAAGTAAATATGATATATTAATCTATGATGAAGATTTTAGCATTTAGTTT 2826
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrrIleIleAspSerVal 121
DB 2827 TGGATTAGAATAAATAAATGGTAAAGTAAATTTACCTGCATATATAATGATAGTGT 2886
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
DB 2887 AAAATAACTCAGGTGGAGTATAGTATTTATTAATTTTATTTAGTATTTTACATTTAA 2946
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAlaPro 161
DB 2947 CAAATGAAGATAGTGAACAAGTATAATTTAGTTATGATATATCAATCAATATGCTCT 3006
QY 162 GlyTyrrAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
DB 3007 GGATAGAATAAATGTTTTTTTGTAACTGTTACTACAAATATGATGGGAAATATGAAGATT 3066
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
DB 3067 TATATAATGGAAATTAATAGATATAATAAGTTAAAGAACTAACTGGAATTAATTT 3126
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
DB 3127 AGCAAACTATACATTTGAAATAAATAAATTTCCAGATACCGGTTTGAATTTACATCAGAT 3186
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAlaLysGluLeuAspGly 241
DB 3187 TCTGATACATCAATATGATGGATAGAGATTTTATATATTTTGTATAGAAATTAGATGCT 3246
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrrThrAsnValLysAspTyrrTrp 261
DB 3247 AAAGATATAATATATTATTAATAGCTTGAATATATACTAATATTTGTAAGAATATTGG 3306
QY 262 GlyAsnAspLeuArgTyrrAsnLysGluTyrrMetValAsnIleAspTyrrLeuAsnArg 281
DB 3307 GGAATGATTTAAGATATAATAAGAATATATATGTTAATATAGATATTATTAATAGA 3366
QY 282 TyrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAspPhe 301
DB 3367 TATATGATGCAACTCAGCAAAATTTTAAATACAGTGAATAATAATATGACTTC 3426
QY 302 AsnGluGlyTyrrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
DB 3427 AATGAGGATATAAATATATAATAAAGAATCAGAGAAATACAAATGATAGAGTA 3486
QY 322 ArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPhe 341
DB 3487 CGAGGAGGAGATATTTATATTTTGATATGACAAATTAATAACAAGCATATAATTTGTT 3546
QY 342 MetLysAsnGluThrMetTyrrAlaAspAsnHisSerThrGluAspIleTyrrAlaIleGly 361
DB 3547 ATGAAGAATGAAGACTATGATGACAGTAATCATAGTACTGAGATATATATGCTAGCT 3606
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
DB 3607 TTAAGAGAACAACAAGGATATAATGATAATATATATTTTCAAAATACAAACAAATGAAT 3666
QY 382 AsnThrTyrrTyrrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
DB 3667 AATATCTTATTTAGGCATCTCAATATTTAAATCAATTTTATGGAGAAATATTTCT 3726
QY 402 GlyIleCysSerIleGlyThrTyrrArgPheArgLeuGlyGlyAspTyrrArgHisAsn 421
DB 3727 GGAATATCTCAATAGGTACTTATCGTTTAGACTTGGAGTGATTTGGTATAGACAAAT 3786
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrrAlaSerLeuLeuGluSerThrSerThr 441
DB 3787 TATTTGGTCCCTACTGTGAACGAAGAAATATGCTTATTTAGAAATCAATCAATCAACT 3846
QY 442 HisTrpGlyPheValProValSerGlu 450
DB 3847 CATTTGGGCTTTTGACCTGTAAAGTGA 3873
RESULT 11
US-10-051-952-4
; Sequence 4, Application US/10051952
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-4
Alignment Scores:
Pred. No.: 7,83e-227 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservatives: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 39 Gaps: 0
US-09-910-186a-10 (1-450) x US-10-051-952-4 (1-3876)
```

QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
DB 2527 ACAATACCCCTTTATATATTTTTCATATACCTAAATATCTTTTATTAAGATATAATAT 2586  
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
DB 2587 GAATATTTTCAATATATATATGATTCATAAATTTTGAGCCTACAAACAGAAAAATACT 2646  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
DB 2647 TTAGTGGATACATCAGGATATATGCGAAGTAGTGAGGAGGAGGATGTCAGCTTAT 2706  
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
DB 2707 CCAATATTTCCATTTGACTTTAAATAGTAGTTTCAGGGAGGAGATAGAGTAAAGTTATA 2766  
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrClnuSerPheSerIleSerPhe 101  
DB 2767 GTACCCAGAAATGAATATATGATATATCTATGATGAAAGTTTATGACATATGTTT 2826  
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
DB 2827 TGGATTAGATAAATAATGCGTAAGTAATTTTACCTGGATATATCTATAATGATAGTGT 2886  
QY 122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
DB 2887 AAAAATAACTCAGGTGGAGTATAGGTATTTAGTAATTTTATGATTTTACTTTAAAA 2946  
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
DB 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATACAAATATGCTCCT 3006  
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
DB 3007 GGATAGAATAAATGCTTTTGTGAAGTGTACTAACAATAATGATGGAATAATGAAGATT 3066  
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
DB 3067 TATATAATGCAAAATTAATAGATACATAAAGTAAAGAACTAAGTGAATTAATTT 3126  
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
DB 3127 AGCAAACTATAAATGTTGAATTAATAAATAATCCAGATACCGGTTTGATTTACTTCAGAT 3186  
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
DB 3187 TCTGATACATCAATATGCGATAGAGATTTTATATATTTCTTAAGAAATAGATGCT 3246  
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAsnValLysAspTyrTrp 261  
DB 3247 AAGAGATATATATATATTTAATAGCTTGAATATATCTAATCTTTGTAAGAGATTATGG 3306  
QY 262 GlyAsnAspLeuArgTyrAspLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
DB 3307 GGAATGATTAAGATATATTAAGAAATATATATGTTTAAATAGATATTTTAATATAGA 3366  
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
DB 3367 TATATGTTGCGAACTCAGCAAAATGTTTAAATACAGTGAATAATAATAGACTTC 3426  
QY 302 AsnGluGlyTyrLysIleIleIleIleArgIleArgGlyAsnThrAsnAspThrArgVal 321  
DB 3427 AATGAAGATATAAATAATATATAAAGAAATCAGAGAAATCAAAATGATAGTATAGTA 3486  
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
DB 3487 CGAGGAGGAGATATTTATATTTTATGATGACATTAATAACAAGCATATATTTGTT 3546  
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
DB 3547 ATGAAGAAATGAACATGATGCGAGATAATCATAGTACTGAAGATATATATGCTATAGT 3606

QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
DB 3607 TTAAGAGAACAAACAAAGGATATAAATGATATATATATATTTCAATACACCAATGAAT 3666  
QY 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
DB 3667 AATACTTTATTTAGGCATCTCAAAATATTTAAATCAAAATTTTAATGGAGAAATATTTCT 3726  
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyLysAspTyrTyrArgHisAsn 421  
DB 3727 GGAATATGTTCAATAGTACTTATCGTTTAGACTTGGAGGTGATTGGTATAGACAACT 3786  
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
DB 3787 TATTTGTCCTACTGTGAAGCAAGAAATATGCTTCATTATTAGAATCAACATCACT 3846  
QY 442 HisTrpGlyPheValProValSerGlu 450  
DB 3847 CATGGGGTTTTGCTACCTGTAAGTGAA 3873  
RESULT 12  
PCT-US97-15394-61  
; Sequence 61, Application PC/TUS9715394  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/15394  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolla, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPND-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1502 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1493  
PCT-US97-15394-61  
Alignment Scores:  
Pred. No.: 8.09e-223 Length: 1502  
Score: 2319.00 Matches: 439  
Percent Similarity: 98.66% Conservative: 3  
Best Local Similarity: 97.99% Mismatches: 5  
Query Match: 96.99% Indels: 1  
DB: 1 Gaps: 0  
US-09-910-186A-10 (1-450) x PCT-US97-15394-61 (1-1502)

```

Qy 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383
Db 1232 AGAACAAACAAGGATATAAATGATATAATTTATTTCAATACAAACCAATGAATAATAC 1291
Qy 383 rTyTyTyTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyI 403
Db 1292 TTATTATTACGCATCTCAAAATATTTAAATCAAAATTTTAAATGCAGAAAAATATTCTGGAAT 1351
Qy 403 eCysSerIleGlyThrTyArgPheArgLeuGlyGlyAspTyrTyArgHisAsnTyLe 423
Db 1352 ATGTTCATAGGACTATCGTTTTAGACTTGGAGTGATGGGTATGATACACAAATATTT 1411
Qy 423 uValProThrValIysGlnGlyAsnTyTyArgAlaSerLeuLeuGluSerThrThrHisTr 443
Db 1412 GGTGCCACTGTGAAGCAAGGAAATATGCTTCATTATTAGAAATCAACATCAACTCATTTG 1471
Qy 443 pGlyPheValProValSerGlu 450
Db 1472 GGGTTTTGTACCTGTAGTGAA 1493

RESULT 13
US-08-704-159-61
; Sequence 61, Application US/08704159
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingclia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..1493
US-08-704-159-61

Alignment Scores:
Pred. No.: 8,09e-223 Length: 1502
Score: 2319.00 Matches: 439
Percent Similarity: 98.66% Conservative: 3
Best Local Similarity: 97.99% Mismatches: 5
Query Match: 96.99% Indels: 1
DB: 11 Gaps: 0

US-09-910-186A-10 (1-450) x US-08-704-159-61 (1-1502)

```



QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTy 23  
Db 152 CATATCGAAGTCTCATATAGCGTACATGGCTTTATTAAACAGATATATTAATGAATA 211  
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43  
Db 212 TTTCAATAATATTAATGATTCACAAATTTTGAGCTTACAAAACAGAAAAATACTTTAGT 271  
QY 43 IAspThrSerGlyTyrAsnAlaGluValSerGluGluLysAspValGlnLeuAsnProI 63  
Db 272 GGATACATCAGATATATATGAGAAGTGAAGAGAGCGATGTTTCAGCTTAATCCAAT 331  
QY 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh 83  
Db 332 ATTTCATTGACTTTTAATTAATAGTAGTTTCAGGGGAGGATAGAGGTAAGTTATAGTAAC 391  
QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheThrPl 103  
Db 392 CCAGAAATGAAATATGTATATAATCTATGATGAAGTTTGTAGCAATAGTTTGGAT 451  
QY 103 eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123  
Db 452 TAGATAATAATAGGGTAAGTAATTTACCTGGATATACATATATGATAGTCTTAATAA 511  
QY 123 rAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs 143  
Db 512 TAACTCAGCTGGAGTATAGTATATATAGTAATTTTGTAGTATTTTACTTTAAACAAA 571  
QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTy 163  
Db 572 TCAAGATAGTGNACAAAGTATAAATTTAGTATGATATATCAAAATAGCTCTCGATA 631  
QY 163 rAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl 183  
Db 632 CAATAAATGGTTTTTGTAACTGTACTAAATATGATGGAATATGGAATATGATATAT 691  
QY 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy 203  
Db 692 AAATGGAAAATTAATAGATATATAAAGTTAAAGAACTAACTGCGAATTAATTTAGCAA 751  
QY 203 sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223  
Db 752 AACTATACAATTGAATAAATAAATAATCCAGATACGGTTTGTACTTACTCAGATTCGA 811  
QY 223 rAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243  
Db 812 TAACATCAATATGTGGAAGAAGATTTTATATATTTGTGTAAGAATATAGATGTAAGA 871  
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAs 263  
Db 872 TATTAATATATTTAATAGCTTGAATATACATTAATGTTGTAAGAATATTTGGGAAA 931  
QY 263 nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMe 283  
Db 932 TGATTTAGATATAATAAAGAATATATATGTTAAATAGATATATTTAAATAGATATAT 991  
QY 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAspPheAsnGl 303  
Db 992 GTATCGAACTACGACAAATGTTTTTAATACACGTAGAAATAATATGCTTCAATGA 1051  
QY 303 uGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323  
Db 1052 AGGATATAAATATATAAATAAAGAATCAGAGGAATACAAATGATAGTACTAGTACGAGG 1111  
QY 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy 343  
Db 1112 AGGAGATATTTATTTTGTATGACAAATTAATAACAAGCATATATTTGTTTATGAA 1171  
QY 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363  
Db 1172 GAATGAAACTATGTATGAGATATATCATAGTACTGAAAGATATATATGCTAGTGTAAAG 1231

QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
Db 1232 AGAACAAACAAAGATATAATGATATAATATATATTTCAATCAACCAATGAATATATAC 1291  
QY 383 rTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIl 403  
Db 1292 TTAUTATTACGATCTCAATATATTAATCAAAATTTAATGAGAAAAATATTTCTGGAAT 1351  
QY 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyrLe 423  
Db 1352 ATGTTCAATAGTACTTATCGTTTACACTTGGAGTGATGTTGATATAGACACAAATTTT 1411  
QY 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHistr 443  
Db 1412 GGTGCTACTGTGAAGCAAGAAATATGCTTCATATTAGAAATCAACATCAACTCATTC 1471  
QY 443 pGlyPheValProValSerGlu 450  
Db 1472 GGGTTTTGTACTGTAAGTAA 1493  
RESULT 14  
US-10-271-012-61  
; Sequence 61, Application US/10271012  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/271.012  
; FILING DATE: 15-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8336  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1502 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1493  
; SEQUENCE DESCRIPTION: SEQ ID NO: 61:  
US-10-271-012-61  
Alignment Scores: 8.09e-223 Length: 1502  
Pred. No.: 2319.00 Matches: 439  
Score: 98.66% Conservative: 3  
Best Local Similarity: 97.99% Mismatches: 5



```
Query Match: 96.99% Indels: 1
DB: 43 Gaps: 0
US-09-910-186a-10 (1-450) x US-10-271-012-61 (1-1502)

QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTy 23
DB 152 CCATATCGAAGGTCGTCATATGCGTAGCATGCTTTATTAAAGATATAAATAATGAATA 211
QY 23 rPheAsnAsnIleAsnAspSerLysIleIleSerLeuGlnAsnArgLysAsnThrLeuVa 43
DB 212 TTTCATATATATATGATTCACAAATTTTGAGCCACAAACACAAATAATCTTAGT 271
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl 63
DB 272 GGATACATCAGGATATAATGCAGAGAGTGCAGTGAAGCGCATGTCAGCTTAATCCAAT 331
QY 63 ePheProPheAspPheLysLeuGlySerSerClycLusArgGlyLysValIleValTh 83
DB 332 ATTTCATTTGACTTTAAATTAGGTAGTTCAGGGAGGATAGAGTAAAGTTATAGTAAC 391
QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTyrIl 103
DB 392 CCAGATGAATAATTTGTATATAATTCATGTATGAAAGTTTACCATAGTTTTCGAT 451
QY 103 eArgIleAsnLysTyrValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123
DB 452 TAGAATAAATAATGAGTGAAGTAAGTAATTTACCGGATATATCTAATGATGTTAAAAA 511
QY 123 nAsnSerGlyTyrPserIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs 143
DB 512 TAACTCAGGTGGATATAGGTATTATTAGTAATTTTATTAGTATTACTTTAAACAAAA 571
QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy 163
DB 572 TGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAATAATGCTCCTGGATA 631
QY 163 rAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl 183
DB 632 CATAAATGGTTTTTGTAACTGTACTTAACAATATGATGGGAATATGAAGATTTATAT 691
QY 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy 203
DB 692 AAATGGAATAATTAAGATACATATAAAGTTTAAAGAACTAACTGGAATTAATTTTGAAC 751
QY 203 sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223
DB 752 AACTATAACATTGAAATAAATAAATAATCCAGATACCGGTTTGATCTTCAGATTCCTGA 811
QY 223 pAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243
DB 812 TAAACATCAATATGTGATTAAGAGATTTTATATATTTCCTAAAGAAATAGATGTAAGA 871
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTyrGlyAs 263
DB 872 TATTATATATTATTAAATAGCTTGCATATATCTAATGTTGTAAAGATTAATGGGAAA 931
QY 263 nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMe 283
DB 932 TGATTTAAGATATAATAAAGATATATATGTTAATATAGATATTATTAATAGATATAT 991
QY 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAspPheAsnGl 303
DB 992 GTATGGCACTCAGCAAAATTTGTTTTAATACAGTGAAGATATAATAGTACTTCAATGA 1051
QY 303 uGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323
DB 1052 AGGATATAAATTAATAAAGATACAGAGAAATACAAATGATAGTACAGGACGAGG 1111
QY 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy 343
DB 1112 AGGAGATATTTTATATTTTGATATGACAAATTAATAACAAGCATATATATTTGTTTGA 1171
```

QY 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363  
DB 1172 GAATGAACACTATGTATGCAGATAATCATAGTACTGAGATATATATGCTATAGGTTAAG 1231

QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPrometAsnAsnTh 383  
DB 1232 AGAACAAACAAAGGATATAAATGATAATATTATATTCAATACAAACCAATGAATAATAC 1291

QY 383 rTyrTyrTyrAlaSerGlnIlePheLysSerAspPheAsnGlyGluAsnIleSerGlyIl 403  
DB 1292 TTATTATACGCATCTCAATATTATTAATCAATTTTATGAGAAATATTTCTGGAAT 1351

QY 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyAspTyrTyrArgHisAsnTyrLe 423  
DB 1352 ARGTTCAATAGTACTTATCGTTTATAGACTTGAGAGTGATGATAGACACAATTTATT 1411

QY 423 uValProThrValLysGlnGlyAspTyrAlaSerLeuLeuGluSerThrSerThrHisTr 443  
DB 1412 GGTGCTACTGTGAAGCAGGAATTTATGCTTCATTATTAGATCAACATCAACTCATTTG 1471

QY 443 pGlyPheValProValSerGlu 450  
DB 1472 GGGTTTGTACCTGTAAGTGAA 1493

RESULT 15  
US-09-611-419A-11  
; Sequence 11, Application US/09611419A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Leonard A.  
; APPLICANT: Byrne, Michael P.  
; APPLICANT: Middlebrook, John L.  
; APPLICANT: Lapenotiere, Hugh  
; APPLICANT: Clayton, Michael A.  
; APPLICANT: Brown, Douglas R.  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; FILE REFERENCE: A33626 067252.0105  
; CURRENT APPLICATION NUMBER: US/09/611.419A  
; CURRENT FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: PCT/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/146,192  
; PRIOR FILING DATE: 1999-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1374  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic construct based on BONTA HC  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10)...(1362)  
US-09-611-419A-11

Alignment Scores:  
Pred. No.: 7,949-77 Length: 1374  
Score: 868.50 Matches: 197  
Percent Similarity: 62.06% Conservative: 86  
Best Local Similarity: 43.20% Mismatches: 145  
Query Match: 36.32% Indels: 28  
DB: 23 Gaps: 11





Db 1120 GTTCCATTGTGATCAATGCGCCTTTACAGCCTCTTTAAAGGCCATAATCCTAATATCTAT 1179  
Qy 14 SerLeuLeuLysAspIleLeuAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeu 33  
Db 1180 ACTTGTGTACAAATGATA-----GATACAAAGG-----AGAGAATATGAATATATA 1227  
Qy 34 SerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAla-----Glu 51  
Db 1228 GTCAGGAAGATTCACAAAGATACTTTGAAGAATAAGAAAGTTTATACAGCAAGAAAGAA 1287  
Qy 52 ValSerGluGluGluAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGly 71  
Db 1288 ATTGAGCAGTATGGTGTGTA----- 1308  
Qy 72 SerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyr--- 90  
Db 1309 -----AACGACATACGATAAATAATAATAATATTTGTATATAG 1350  
Qy 91 -----AsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLys 107  
Db 1351 CATGCCGGAATAGGAAATACATCTTTTGGTGATAAATATTTATTTATGAAATAGAG 1410  
Qy 108 TrpValSerAsnLeuProGlyTyrThrIleAspSerValLysAsnAsnSerGlyTyr 127  
Db 1411 CAGAGTATT-----ATGAGTATATAAGTCAATTAACATATGTCAAATGAA 1458  
Qy 128 SerIleGlyIleLeuSerAsnPheLeuValPheThrLeu----- 140  
Db 1459 AGCGTAGAGGCTATTCAAAATCAGTAGAGTTTGGCTAAATATGAATATACATCGTTG 1518  
Qy 141 -----LysGlnAsnGluAsp-----SerGluGlnSerIleAsnPhe 152  
Db 1519 GATTGTAAAGAAAGAGGAATACCTTTAAGATGTTTCAATATAGCAAAATATCGATG 1578  
Qy 153 SerTyrAspIleSerAsnAsnAlaProGlyTyrAsnLysTrpPheValThrValThr 172  
Db 1579 GTATAC-----GGAGCTGTGTGGACAGGGAATCACTTTAATAAACCATAT 1626  
Qy 173 AsnAsnMetMetGlyAsnMetLysIleTyrIle-----AsnGlyLysLeuIleAsp 189  
Db 1627 TGCAATTTTCTATGATTAAGATGTTATCGTATAGCAAAATCTAATCTCGAGTAGAC 1686  
Qy 180 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209  
Db 1687 AATATTAAGAAAA-----ATCAAAATATCTAACATTAACATTAACGCTCTACTATT 1734  
Qy 210 AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle 229  
Db 1735 TCTAAA-----TTTTATATATGATAAGAAAGATGACTGTGTTAATA 1779  
Qy 230 ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn 249  
Db 1780 ATCGAT-----GAGGCAGGTACAGTTAGTAATAAGGACATGAATCGAATCTGTGAA 1830  
Qy 250 SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLys 269  
Db 1831 ACAAGCAATTGATATATA-----TTAATGTGCGGTAT 1866  
Qy 270 GlutTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsn----- 286  
Db 1867 AATTATCAATCGATCGATAGATTTTGGAAUTTGTTGCGAAATGCCAAGAGATGTTTG 1926  
Qy 287 SerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLys 306  
Db 1927 TCAAAAAATAAATC-----AACGAATCACTATGATGATAT--- 1962  
Qy 307 IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIle 326  
Db 1963 -----CGAATCAAAATATGATGATTTA 1983  
Qy 327 LeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThr 346  
Db 1984 CTTTACTTTTGGAAATCTGTTAGAAAAAAGAGATAATTA-----AATGAAT 2034

Qy 347 MetTyrAlaAspAsnHisSerThr----- 354  
Db 2035 ATTAATATGAATAAATATCTACAGATAGTAGAAGATATATTTAATGAATCAATAAA 2094  
Qy 355 -----GluAspIleTyrAlaIleGlyLeuArgGlu--- 364  
Db 2095 GATCAAAATATCTTGTGTTTAAATACGATGTATATATGGTATCAATAAACATTAAGA 2154  
Qy 365 -----GlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro----- 379  
Db 2155 TTATTACAAGCAAAATATAAAATGATCTGTAATTTGGGTGTGAAGAATATAAAGTT 2214  
Qy 380 -----MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsn 394  
Db 2215 GGTGATCCTATCTATTTATGAAGAACTAATAAGTACTACCAACTACTTTT---AATAAT 2271  
Qy 395 PheAsnGlyGluAsnIle 400  
Db 2272 TTAAGAGGTCATAAAT 2289

## RESULT 2

US-10-092-411A-879

; Sequence 879, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/10/092,411A

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 879

; LENGTH: 4008

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-879

## Alignment Scores:

Pred. No.:	0.000521	Length:	4008
Score:	121.50	Matches:	97
Percent Similarity:	34.83%	Conservative:	74
Best Local Similarity:	19.76%	Mismatches:	193
Query Match:	5.08%	Indels:	127
DB:	6	Gaps:	23

US-09-910-186A-10 (1-450) x US-10-092-411A-879 (1-4008)

Qy 10 TyrThrAsnAsnSerLeuLeuLysAspIleLeuAsnGluTyrPheAsnAsnIleAsnAsp 29  
Db 2698 TACCTAACCGTATCTCCCT-----ATCAATPAACACGCTACTCAAGGTATATAC 2748  
Qy 30 SerLysIleSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsn 49  
Db 2749 ACGTATGTACTTCTCAAGATGGACAAAGTAATACTCCATTA----- 2790  
Qy 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLys 69  
Db 2791 -----GGATGGGTAAACATTAATGATGTGACAACTCAAAAT----- 2826  
Qy 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89  
Db 2827 ATCGAAAAACAACTCAGTCTATAGTAAATATTTCAGTAAACCTACAAATATGTCCTA 2886  
Qy 90 TyrAsnSerMetTyr----- 94  
Db 2887 TATTCTTGTCTGGGTGCTATAAAACCAACAATTAAGTACACCTAATACGCTAGCTAAT 2946

QY	95	GlucosePheserPheThrIleArgLysAsnSerGlyTyrPheValSerAsnLeuProGly	114	QY	415	GlyAspTyrPheArgHisAsn	421
Db	2947	CAAGCATTTAATGCTCCAAAGCTGTTACGTTGGTAAA	2988	Db	3913	ATTAATTTGGCAAAATGTCACAGCCACAAAGTATTTAAACAATTTATGATAAAATTTGCTGAA	3972
QY	115	TyrThrIleLeuAspSerValLysAsnSerGlyTyrPheSerIleGlyIleIleSerAsn	134	QY	422	-----TyrLeuValProThrValLys	428
Db	2989	TTATATCTATACGGTACAGTCAATACAGACAGATGG	3039	Db	3973	GTCGTAAGTACTTCCGAATTCACACATACAAA	4005
QY	135	PheLeuValPheThrLeuLysGlnAsnGluAspSerGluInSerIleAsnPheSerTyr	154	RESULT 3			
Db	3040	GATTTAATC-----CAAAACAGTACTGACGCTCAATCAACCAATATAACTAT	3087	US-10-092-411A-2477			
QY	155	Asp---IleSerAsnAlaProGlyTyrAsnLysTyrPhePheValThrValThrAsn	173	US-10-092-411A-2477			
Db	3088	ACTTTTGTATCAATATAGTAAAGTTAT-----TTCATATGATGATCCCAACAAA	3138	GENERAL INFORMATION:			
QY	174	AsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysVal	193	APPLICANT: Lynn Doucette-Stamm et al			
Db	3139	GCAAACCGATATCTTTAAACCAATATATGAA---CAAACITTCACAGTCAATTAAGCAA	3195	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS			
QY	194	LysLeuLeuThrGlyIleAsnPhe	201	TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
Db	3196	AAAAATTAATGGCTTAATGCTACTATGCTCAACTTTAGCGGTAATAATGTTGG	3255	FILE REFERENCE: 032796-101			
QY	202	-----SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr	219	CURRENT APPLICATION NUMBER: US/10/092,411A			
Db	3256	ATAAAATCAACTGACTTAGTTAAGGAAAAATTAATATGATATGATACGTGAATGACITTA	3315	CURRENT FILING DATE: 2002-03-07			
QY	220	SerAspSerAspAsnIleAsn	226	PRIOR FILING DATE: 1998-08-13			
Db	3316	AATAACGCGATAAATTCACATCTGCTTAAATATAAACCACAAAGTACAAAATGAGCCT	3375	PRIOR FILING DATE: 1997-11-08			
QY	227	-----MetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAsn	245	PRIOR APPLICATION NUMBER: US 09/134,001			
Db	3376	TTGAAATGGTCAATGCTAATATATGCTCAAAATTA---AATGCTATGATGATACAAAG	3429	PRIOR FILING DATE: 1997-08-14			
QY	246	IleLeuPheAsn-----SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsn	263	NUMBER OF SEQ ID NOS: 5676			
Db	3430	CGTTTACTGATTCATCCCTTAAATAT-----CAA	3462	SEQ ID NO 2477			
QY	264	AspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMet	283	LENGTH: 2793			
Db	3463	TTCTTACGTTTATGATCAACCAACAAATCTG---TCAGCACAGCTCTCAATTAATTTA	3519	TYPE: DNA			
QY	284	-----TyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn	299	ORGANISM: Staphylococcus epidermidis			
Db	3520	AAAGGCAAGGTGCTTGAACACCAAGCGCTGCTATTTAGCCACAGCTGCAGTAAGTAT	3579	US-10-092-411A-2477			
QY	300	AspPheAsnGluGlyTyrLysIleIleLysArgIleArg---GlyAsnThrAsnAsp	318	Alignment Scores:			
Db	3580	GGTCTAATGAAATTTATCTATCTACATGCTTTAGTAGAACAAGTAAATGGACITCA	3639	Pred. No.:			
QY	319	ThrArgValArgGlyLysPheLeu-----TyrPheAspMetThrIleAsnAsnLys	336	Score:			
Db	3640	CAACTTCTTAAGGTGGAGATGTTTCAAAAGGTAAATTCACAACTAAAACAGGTCACAAA	3699	Matches:			
QY	337	AlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp	356	Percent Similarity:			
Db	3700	TACCATTAATGCTTTGGAATGGTCA-----TTTGACAAATTAATGCACTGTAGAT	3750	Conservative:			
QY	357	IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln	376	Best Local Similarity:			
Db	3751	GGTATCAATACGCTAAAATGCTGATGGAATCTGCTCTTAAAGCAATTAATGGTGGC	3810	Query Match:			
QY	377	IleGlnProMetAsnAsnThrTyrTyrAlaSerGln-----IlePheLysSerAsn	394	DB:			
Db	3811	GCTAAATCAATGGAATTCATACGTAAGGACGACGACAAATACGCTATATAAAATGCGT	3870	US-09-910-186A-10 (1-450) x US-10-092-411A-2477 (1-2793)			
QY	395	PheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGly	414	QY	9	SerTyrThrAsn-----AsnSerLeuLeuLysAspIleIle	20
Db	3871	TGGAAATCGTCAAAAC-----CCTGGTACGCATCAATATGCAACTGAT	3912	Db	670	AGCTATAAAATATAGATGAAATTTCAAAATCAAGATGATTTAAATTTACCAATA	729



```
; Sequence 1411, Application US/10240425
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Bolland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1411
; LENGTH: 2376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. U02020
US-10-240-425-1411

Alignment Scores:
Pred. Score:      0.000907      Length:      2376
Score:            116.50        Matches:     77
Percent Similarity: 36.17%      Mismatches: 59
Best Local Similarity: 20.48%   Indels:    107
Query Match:       4.87%       Gaps:       18
DB:                6

US-09-910-186A-10 (1-450) x US-10-240-425-1411 (1-2376)

Qy  7 IlePheSerTyr-----ThrAsnAsnSerLeuLeuLysAsp--- 18
      :::::|||||
Db 124 GTTTATTCCTACTTTGAATGCGGTGAAAGACAGACAAGAACTCCAATTAAAGAAGGTG 183
Qy  19 -----IleLeuAsnGluTyrrPhe----- 24
      :::::|||||
Db 184 AAATATGAGGAACAGTATTATTGGTTGCCAGTCACTCTTATAAGTACTTAAGAGGT 243
Qy  25 -----AsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLys 39
      ||| :::::|||
Db 244 AAGTAGTAACCAAGAGAAAATCCAGGAGGCCAAGATGTCTACAAGAACACATTTCCAA 303
Qy  40 AsnTrrLeuValAspThrSerGlyTyrrAsnAlaGluValSerGluGluGlyAspValGln 59
      :::::|:::|
Db 304 GATGATGCTTTAATGAAAAGGATGSAAC--TACATTCTTGAGAGTATGATGGGCAT 360
Qy  60 LeuAsnProIlePheProPheAspPheLysLeuGlySerSerGly-----GluAspArg 77
      ||| :::::|||
Db 361 CTW-----CCAAATAGAAAATAAAGCTGTTCCTGAGGGCTTTGTTCATTCCCCAGA 408
Qy  78 GlyLysValIleValThrGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPhe 97
      ||| :::::|
Db 409 GGAATGTCTCTTCACG-----GTGGAACACACAGACATCCAGAGTGTTCAC 453
Qy  98 SerIleSerPheTrpIleArg-----IleAsnLysTrpValSerAsnLeuProGly 114
      :::::|||||
Db 454 TGCGTTACAAATTTGATTTGAGCAGATTCCTTTGTCAGTCCCTGG----- 495
Qy  115 TyrThrIleIleAspSerValLysAsnAsnSerSerglyTrpSerIleGlyIleIleSerAsn 134
      ||| :::::|
Db 496 TATCCAATC-----ACAGTGGCCACAAAAATCTCTAGAGACGACGAAGAAAATATTGGCCAAA 549
Qy  135 PheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIle---AspPheSer 153
      :::::|||||
Db 550 TATTTCCTAGAAACCTTCGGTAACTTATGATGGTCTCGGAATACAAAGTTACATGATTTCGC 609
```

```
Qy 154 TyrAsp-----IleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheValThrVal 171
|||
Db 610 TACAGAGAGTCTCTCCCAAGAGACTGCTGGC-----642
Qy 172 ThrAsnAsnMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIle 191
|||
Db 643 -----ATAGAGGATCTGCTCACTTGGTTAACTTCAAGAGACACATACAGTA 690
Qy 192 LysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLys 211
|||
Db 691 -----GCAGAGACTTCTCTTAATAAAAAATATTATGAACGAGAAAGATCCT 735
Qy 212 IleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgasp 231
|||
Db 736 GTTCAGGCTATTCGTCCAGCAGCAACACAGTACCATAACACCTTGGGGAAAGAC 795
Qy 232 -----PheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIle 246
|||
Db 796 CATGAAAAAGATGCTTTGAACATATTGTAACACAGTTCATCGCTGTATCTGTG 855
Qy 247 LeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArg 266
|||
Db 856 GTACCGATACGTATGACATTTATATGCGGTGAGAAATATGGGTGAAGATCTAAGA 915
Qy 267 TyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsn 286
|||
Db 916 CAT-----TTAATAGTATCGAGA 933
Qy 287 SerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLys 306
|||
Db 934 AGTACACAGCA-----CCA 948
Qy 307 IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIle 326
|||
Db 949 CTAAATACAGACCTGATCTCGAACCCTCTTGACACTGCTTAAGGTTTGGAGATT 1008
Qy 327 Leu-----TyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
|||
Db 1009 TTAGTAAAGATTCTCTGTACTGAGAAGCTCAAGGGTTACAAGTTG 1056
```

## RESULT 6

```
US-10-092-411A-405
; Sequence 405, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 405
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-405
```

```
Alignment Scores:
Pred.No.: 0.00153 Length: 3456
Score: 116.50 Matches: 97
Percent Similarity: 34.44% Conservative: 90
Best Local Similarity: 17.86% Mismatches: 175
Query Match: 4.87% Indels: 181
DB: 6 Gaps: 27
```

US-09-910-186a-10 (1-450) x US-10-092-411A-405 (1-3456)

```
Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
|||
Db 214 CTACCAAGTTCTCCAGTACCACACAGCAGCAATACAGCTATCTGTTTCAGTACCAAGAA 273
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
|||
Db 274 -----GACGAAATCACTGCCAATCTTACGCAACAACAATATACAAA 315
Qy 43 Val-----AspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVal 58
|||
Db 316 TTAACAAAGTACACAGATGTTAAACAAATGAATCGACAAATGTCTAGTAAGTCGATA 375
Qy 59 GlnLeuAsnProIlePhePheProPheAsp---PheLysLeuGlySerSerGlyGluAspArg 77
|||
Db 376 GAATTAAT-----TTTCCATCAACAATTTAATCTAGGGTTTACATTTCAAGATACA 429
Qy 78 GlyLysVal-----Lys 83
|||
Db 430 CCGGGTGTGTACTCTAAGTTCGAACGATCAGTCAGTACAGACGAATTTATGTATACA 489
Qy 84 GlnAsn-----GluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
|||
Db 490 AGTAACCTGTTGTTTATACAGTAGACTATATCATGTCCAAATCAGCGTTGACITTA 549
Qy 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
|||
Db 550 TTTATGAACGCTATTAAATGAA-----GTTGGTATACCAATTAATTTTGTGATTAATCAG 603
Qy 121 Val-----Lys 122
|||
Db 604 ATTGATAAACATAATGAAGAAGAAATACATTGTGAACCTTTAAATCAAGATCGAAAA 563
Qy 123 AsnAsnSerGlyTrpSerIleGlyIle-----131
|||
Db 664 TCAATCAAGACTCGGATATCAAACTTCAAGATACTTAATTCGTTTCAAAAGTTGATCAT 723
Qy 132 -----IleSerAspPheLeuValPheThrLeuLysGlnAsnGlu 144
|||
Db 724 CCACAGATGAATGTAGAACTTCAAACTTTCTAGTATTTATGGATCAACATCGTGA 783
Qy 145 AspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyrAsn 164
|||
Db 784 TCAACAGAGACTATGTTAATAGAACAAATCAATTCATTCACGACGCACAA-----834
Qy 155 LysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsn 184
|||
Db 835 -----TACATATACATCAAAATGAATGCA-----861
Qy 185 GlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThr 204
|||
Db 862 --TCTATTCTTGACACCTTCAAAATTAATGAA-----GAACAATTCGAGGAAGCA 909
Qy 205 ---IleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsp 223
|||
Db 910 TATATTCAAATTTCAACAAATCAAA-----GAAGTCAGCGCAGAGACACAA 954
Qy 224 AsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAsp 243
|||
Db 955 TTGCTCAAT-----GAC 966
Qy 244 IleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsn 263
|||
Db 967 TCTAATCAATATTATTAATTAATTAATAAAGAGCGT-----AAAGATATATTAGATAAT 1020
Qy 264 -----AspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle---276
|||
Db 1021 GCTTATATCATGACGTACGATATCGCGTAATCTTACGGATTAATTAGAACATCGGCA 1080
Qy 277 ---AspTyr---LeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThr 294
|||
Db 1081 ACTGATTTTAAAGTGAATGGATT-----TTTAATAAA 1113
Qy 295 ArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGly 314
```



```

Db 1114 AGGAGAAAAAGAGAGACAA-----ATCAACGACTTAATGAG 1155
Qy 315 AsnThrAsnAspThrArgValArgGly-----Asp 325
Db 1156 CGGACCACTCAATTCGAAGAAAGTAATCAACAAGTACGACACCACTTCGTGAAGAT 1215
Qy 326 IleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlu 345
Db 1216 ATGTCATTTTAACTAGATCAATAATAAACATCGCTGGAATGAAAAAATATAATCA 1275
Qy 346 -----ThrMetTyrAlaAspAsnHisSerThrGlu 355
Db 1276 GAATATGACGTCGTCGTCCTCACTTATATCATGAGCTATATCAAACTCAACAGCATAGC 1335
Qy 356 AspIleTyrAlaIleGlyLeuArgGluGlnThr---LysAspIleAsnAspAsnIleIle 374
Db 1336 AACACATACGTTTAAACATTTTCAGATGAAGTATTAAGCTTTGAATAAAAATAGAA 1395
Qy 375 PheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsn 394
Db 1396 AATGAGTCAACACCA-----CTATTGAAGAAGCT 1425
Qy 395 PheAsnGlyGluAsnIleSerGlyIleCysSer----- 405
Db 1426 GTCAATCATGATACAAAGTTAATGAATATCGAGTGATCAAAATGAAGATAGCATGAATAT 1485
Qy 406 -----IleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHis 420
Db 1486 GATAGATACATTTGAACATTACACATTAAGGATTCGCTTACATCCACACATACAAACAT 1545
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1546 TACTATATFC-----CATTTAGACGATCTTTAGATAAATTAATTTGAAGAACAGAG 1596
Qy 441 ThrHisTrp 443
Db 1597 ACTCATTTT 1605

```

## RESULT 7

```

US-10-092-411A-2277
; Sequence 2277, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 2277
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-2277

```

## Alignment Scores:

```

Pred. No.: 0.000794 Length: 1494
Score: 114.50 Matches: 90
Percent Similarity: 35.27% Conservative: 56
Best Local Similarity: 21.74% Mismatches: 158
Query Match: 4.79% Indels: 110
Dbs: 6 Gaps: 21

```

US-09-910-186a-10 (1-450) x US-10-092-411A-2277 (1-1494)

Qy 10 TyrThrAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAsp 29

```

Db 300 TACACGTTTAAACGGCATGAGATGAATAATCAATGAATTCAGTCGACATAGCAACGAT 359
Qy 30 SerLysIleLeuSerLeuGlnAsnArgLys-----Asn 40
Db 360 TGTACACACGTTGCCCTTAGAATTAGAAGGGTTTAATTTATATCATTTTAAATGATTAAAC 419
Qy 41 ThrLeu-----Val-AspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVa 58
Db 420 ACAATTAGATAACATGAAGTTTCAGGTTACAAA-----TT 455
Qy 58 LglnLeuAsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGl 78
Db 456 TAATTAGAAATCGATTGCATCAATTAATGTAGCTGTGAACGACGGACCTCAAAAAGC 515
Qy 78 YLysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSe 98
Db 516 TGTCCCTCAACAGGTTTAATTAATCATTTAGCCAGTGTAAAGCTGTAAACAAGTTAGG 575
Qy 98 rIleSerPheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIl 118
Db 576 ATTCGAA-----CAAAATACTGTGTGGCTTTCGGTCTTACCTATATATCATATT-- 624
Qy 118 eAspSerValLysAsnAsnSerGlyTrpSerIle-----GlyIleIleSerAsnPh 135
Db 625 -----TCTGGCTCAGTGTATTTTCCGCGCAGTGATGATCAAGGAT 665
Qy 135 eLeuValPheThrLeuLys-----GlnAsnGluAspSerGluGlnSerIleAsnPheSe 153
Db 666 CACTGTGACAGACTGTTTAAAAAGTTTCAAACTGATGATATGTTAAACACAAATAAAG---AC 722
Qy 153 rTyrAspIleSerAsnAsnAla-----ProGlyTyrAsnLysTrpPhePheValThrVa 171
Db 723 TTATCCCAATCACCATATGTCCTTGTCCCAACAACTTAAGTGTGTTAATGGATCGCAG 782
Qy 171 lThrAsnAsnMetMetGlyAsnMetLysIleIleAsnGly-----Iy 186
Db 783 ATTGACTCAACCAATTTCTTTAGAAAAAATTCGTAGGTGGTCTAAATATTACACCACA 842
Qy 186 sLeuIleAsp-----ThrIleLysValLysGluLeuThrGlyIleAsnPheSerly 203
Db 843 ATTAATTGACCAAGCATGATGCTTACCTGCTGATATATAATCTTTGGTATGACAGA 902
Qy 203 sThrIleThrPheGluIleAsnLysIlePro-----AspThrGl 216
Db 903 AACTGTCTCTCAGTTTCTAACAGCCCTCACCTCAATGCTCAAGAAGACGTTTCGATACT-- 960
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAl 236
Db 961 ----GTTGGAAAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1016
Qy 236 aLys---GluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAs 255
Db 1017 ACATGGAGAGCTTATTAATTAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1073
Qy 255 nValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValas 275
Db 1074 ATATTAAAAAGACACATTTGATATGATGAT---GGGTATTTTCAAACTGGACATATGCTGA 1130
Qy 275 nIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrAr 295
Db 1131 AATAGAT----- 1137
Qy 295 gArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAs 315
Db 1138 -----GATGAAGGTTTACGTCATAAATATATATGATCGG----- 1167
Qy 315 nThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAs 335
Db 1168 -CGCAAGAGATTTGATTAAGTGGTGGAGAGATATTTAT----- 1206
Qy 335 nLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGl 355

```

Db	3461	ACCAACATCATCCTTAATACTAAACCTTAACAAAAACACACACAAAAAATAATTTCCAA	3400
QY	96	SerPheSerIleSerPheThrIleArgIleAsnLysTrpValSerAsnLeuProGlyTyr	115
Db	3401	CCAAATTCGCTA-----ATAAACATCATTAACGAAAAATCTCAATAAAATACGACAA	3351
QY	116	ThrIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhe	135
Db	3350	ACTAAATCCACACACACATCAAAAACCTTATCAACCCACATCAAAATTAACCTCATCC	3294
QY	136	LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAsp	155
Db	3293	-----CTAAATACAAAATAATTC-AACATACACAAATCAATAAACAAATACCATTC	3241
QY	156	IleSer-----AsnAsnAlaProGlyTyr	163
Db	3240	GTAAACAAACCAATCAAAAAACCGTAATATTACTCAATAAATACAAAAAACCTTA	3181
QY	164	AsnLys-----TrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
Db	3180	ATAAAATTCAACACCCCTTCATCTAAACACTCAATAAATCAATAATTAATAAAACG	3121
QY	182	TyrIleAsnGlyLysIleuIleAspThrIleLysValLysGluLeuThrGly-----Ile	199
Db	3120	TATCTC-----AAATATATAAACTATTATTATACAAACCCACCAACCAATACATACTA	3067
QY	200	AsnPheSerLysThrIleThrPheGluIle-----Asn-LysIleProAspThrGlyLe	217
Db	3066	ATAATAACAAAACTATAAACCGTTCCTTTAAATAACTAACACAAACAAATATACCTCTCT	3007
QY	217	uIleThrSerAspSerAspAsnIle-----AsnMetTrpIleArgAs	231
Db	3006	CACACTCTCTATTCAACATATATTAATAAATTTCAACCAACATCAACAAAAAATAA	2947
QY	231	pPheTyrIlePheAlaGlyGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLe	251
Db	2946	AAATAAAATATATTCAAATAAA-----AAAAAATAACTCAAAATTAATATACTA	2901
QY	251	uGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeu-----	265
Db	2900	-----TTTACAATAACATAATTATATATTAATAAATCCATTAATCTCAACCCAAATCT	2845
QY	266	-----ArgTyrAsnLys---GluTyrTyrMetValAsnIleAspTyrLe	279
Db	2844	CCTTAACCTAATCAACAACCTTCAACAAAATCTCAAAATTCAAAATCAATATAAAAAATC	2785
QY	279	uAsnArgTyrMetTyrFalaAsnSerArgGlnIle-----ValPheAsn-----	293
Db	2784	ACAAAATTCCTATACACAATATAATAACACAAAAACCAATCATAAATAACTCCC	2725
QY	294	-----ThraGArgAsnAs	298
Db	2724	ATTACGATTACTACAAAAAATAAAATCACTAAAAATCCAAATCAAAAAAATATAA	2655
QY	298	nAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAs	318
Db	2664	AAACCTATTCAAAAAAATCTACAAACCACTACTCTCAAAAAATAAAAAAACAATAA	2605
QY	318	pThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTy	338
Db	2604	-----ATAAAAAACATTCCTACTCATTAATAAAAA-----	2572
QY	338	rAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-Ile	358
Db	2571	-AATCAATATCATAAAAAT-----AACCATACTACCCAAAAATAATTT	2531
QY	358	yrAlaIleGly-----LeuArgLysGlnThrAspIleAsnA	371
Db	2530	ATAAATTCAAATACTATCCCATCAAACTACTACTACTACTTTTTCRAAAATTAATAA	2471
QY	371	spAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSer-----	388
Db	2470	ACTACTTTAAATTCATATAAAACCAAAAAAATACTTATATACCAAAACATCTCAAC	2411

US-10-240-485-131/c  
; Sequence 131, Application US/10240485

; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian

```

; TITLE OF INVENTION: Diagnosis of Diseases Associated with
;
; TITLE OF INVENTION: Metastasis
;

```

: CURRENT FILING DATE: 2002-10-02  
 : CURRENT APPLICATION NUMBER: US/10/240,485  
 : PRIOR APPLICATION NUMBER: PCT/JP01/03070

/ PRIOR FILING DATE: 2001-04-00  
 ; PRIOR APPLICATION NUMBER: DE 10019058.8  
 / PRIOR FILING DATE: 2000-04-06

; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR APPLICATION NUMBER: DE 10043825.1  
; PRIOR FILING DATE: 2000-09-01

```

; SEQ ID NO 131
; LENGTH: 6070
; TYPE: DNA

```

ORGANISM: ALLICIOUS SEQUENCE  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

Alignment Scores:

Score:	112.00	Matches:	113
Percent Similarity:	35.84%	Conservative:	68

Query Match:	4.68%	Indels:	153
DB:	6	Gaps:	26

US-09-910-188A-10 (1-430) X US-10-240-483-131 (1-6070)	
0v	10 TvrHrAsnAsnSeriententIvAsntle-----lleasngltvrbheasno----- 25

Db 3740 TACACAAATATA-----CTAAAAATCTAAAAATAAAFAAATTCCTTAAACACATAC 3687

Db 3686 ACCCTCCCAAACTAAACCAAAAAAATCAATCCCTA-----AATAACTA 3639

Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuasnPro 62

Qy	63	-----IlePheProPhe	66
		.....	

[illegible]

D<sub>b</sub>    3521   CTTCGAAAACATATTCCAAAAAATAAAAAAAAAAACTCCTCCTCACTCATTTTATATA    3462

1.  $\frac{1}{2}$  2.  $\frac{1}{3}$  3.  $\frac{1}{4}$  4.  $\frac{1}{5}$  5.  $\frac{1}{6}$  6.  $\frac{1}{7}$  7.  $\frac{1}{8}$  8.  $\frac{1}{9}$  9.  $\frac{1}{10}$  10.  $\frac{1}{11}$  11.  $\frac{1}{12}$  12.  $\frac{1}{13}$  13.  $\frac{1}{14}$  14.  $\frac{1}{15}$  15.  $\frac{1}{16}$  16.  $\frac{1}{17}$  17.  $\frac{1}{18}$  18.  $\frac{1}{19}$  19.  $\frac{1}{20}$  20.  $\frac{1}{21}$  21.  $\frac{1}{22}$  22.  $\frac{1}{23}$  23.  $\frac{1}{24}$  24.  $\frac{1}{25}$  25.  $\frac{1}{26}$  26.  $\frac{1}{27}$  27.  $\frac{1}{28}$  28.  $\frac{1}{29}$  29.  $\frac{1}{30}$  30.  $\frac{1}{31}$  31.  $\frac{1}{32}$  32.  $\frac{1}{33}$  33.  $\frac{1}{34}$  34.  $\frac{1}{35}$  35.  $\frac{1}{36}$  36.  $\frac{1}{37}$  37.  $\frac{1}{38}$  38.  $\frac{1}{39}$  39.  $\frac{1}{40}$  40.  $\frac{1}{41}$  41.  $\frac{1}{42}$  42.  $\frac{1}{43}$  43.  $\frac{1}{44}$  44.  $\frac{1}{45}$  45.  $\frac{1}{46}$  46.  $\frac{1}{47}$  47.  $\frac{1}{48}$  48.  $\frac{1}{49}$  49.  $\frac{1}{50}$  50.  $\frac{1}{51}$  51.  $\frac{1}{52}$  52.  $\frac{1}{53}$  53.  $\frac{1}{54}$  54.  $\frac{1}{55}$  55.  $\frac{1}{56}$  56.  $\frac{1}{57}$  57.  $\frac{1}{58}$  58.  $\frac{1}{59}$  59.  $\frac{1}{60}$  60.  $\frac{1}{61}$  61.  $\frac{1}{62}$  62.  $\frac{1}{63}$  63.  $\frac{1}{64}$  64.  $\frac{1}{65}$  65.  $\frac{1}{66}$  66.  $\frac{1}{67}$  67.  $\frac{1}{68}$  68.  $\frac{1}{69}$  69.  $\frac{1}{70}$  70.  $\frac{1}{71}$  71.  $\frac{1}{72}$  72.  $\frac{1}{73}$  73.  $\frac{1}{74}$  74.  $\frac{1}{75}$  75.  $\frac{1}{76}$  76.  $\frac{1}{77}$  77.  $\frac{1}{78}$  78.  $\frac{1}{79}$  79.  $\frac{1}{80}$  80.  $\frac{1}{81}$  81.  $\frac{1}{82}$  82.  $\frac{1}{83}$  83.  $\frac{1}{84}$  84.  $\frac{1}{85}$  85.  $\frac{1}{86}$  86.  $\frac{1}{87}$  87.  $\frac{1}{88}$  88.  $\frac{1}{89}$  89.  $\frac{1}{90}$  90.  $\frac{1}{91}$  91.  $\frac{1}{92}$  92.  $\frac{1}{93}$  93.  $\frac{1}{94}$  94.  $\frac{1}{95}$  95.  $\frac{1}{96}$  96.  $\frac{1}{97}$  97.  $\frac{1}{98}$  98.  $\frac{1}{99}$  99.  $\frac{1}{100}$  100.  $\frac{1}{101}$  101.  $\frac{1}{102}$  102.  $\frac{1}{103}$  103.  $\frac{1}{104}$  104.  $\frac{1}{105}$  105.  $\frac{1}{106}$  106.  $\frac{1}{107}$  107.  $\frac{1}{108}$  108.  $\frac{1}{109}$  109.  $\frac{1}{110}$  110.  $\frac{1}{111}$  111.  $\frac{1}{112}$  112.  $\frac{1}{113}$  113.  $\frac{1}{114}$  114.  $\frac{1}{115}$  115.  $\frac{1}{116}$  116.  $\frac{1}{117}$  117.  $\frac{1}{118}$  118.  $\frac{1}{119}$  119.  $\frac{1}{120}$  120.  $\frac{1}{121}$  121.  $\frac{1}{122}$  122.  $\frac{1}{123}$  123.  $\frac{1}{124}$  124.  $\frac{1}{125}$  125.  $\frac{1}{126}$  126.  $\frac{1}{127}$  127.  $\frac{1}{128}$  128.  $\frac{1}{129}$  129.  $\frac{1}{130}$  130.  $\frac{1}{131}$  131.  $\frac{1}{132}$  132.  $\frac{1}{133}$  133.  $\frac{1}{134}$  134.  $\frac{1}{135}$  135.  $\frac{1}{136}$  136.  $\frac{1}{137}$  137.  $\frac{1}{138}$  138.  $\frac{1}{139}$  139.  $\frac{1}{140}$  140.  $\frac{1}{141}$  141.  $\frac{1}{142}$  142.  $\frac{1}{143}$  143.  $\frac{1}{144}$  144.  $\frac{1}{145}$  145.  $\frac{1}{146}$  146.  $\frac{1}{147}$  147.  $\frac{1}{148}$  148.  $\frac{1}{149}$  149.  $\frac{1}{150}$  150.  $\frac{1}{151}$  151.  $\frac{1}{152}$  152.  $\frac{1}{153}$  153.  $\frac{1}{154}$  154.  $\frac{1}{155}$  155.  $\frac{1}{156}$  156.  $\frac{1}{157}$  157.  $\frac{1}{158}$  158.  $\frac{1}{159}$  159.  $\frac{1}{160}$  160.  $\frac{1}{161}$  161.  $\frac{1}{162}$  162.  $\frac{1}{163}$  163.  $\frac{1}{164}$  164.  $\frac{1}{165}$  165.  $\frac{1}{166}$  166.  $\frac{1}{167}$  167.  $\frac{1}{168}$  168.  $\frac{1}{169}$  169.  $\frac{1}{170}$  170.  $\frac{1}{171}$  171.  $\frac{1}{172}$  172.  $\frac{1}{173}$  173.  $\frac{1}{174}$  174.  $\frac{1}{175}$  175.  $\frac{1}{176}$  176.  $\frac{1}{177}$  177.  $\frac{1}{178}$  178.  $\frac{1}{179}$  179.  $\frac{1}{180}$  180.  $\frac{1}{181}$  181.  $\frac{1}{182}$  182.  $\frac{1}{183}$  183.  $\frac{1}{184}$  184.  $\frac{1}{185}$  185.  $\frac{1}{186}$  186.  $\frac{1}{187}$  187.  $\frac{1}{188}$  188.  $\frac{1}{189}$  189.  $\frac{1}{190}$  190.  $\frac{1}{191}$  191.  $\frac{1}{192}$  192.  $\frac{1}{193}$  193.  $\frac{1}{194}$  194.  $\frac{1}{195}$  195.  $\frac{1}{196}$  196.  $\frac{1}{197}$  197.  $\frac{1}{198}$  198.  $\frac{1}{199}$  199.  $\frac{1}{200}$  200.  $\frac{1}{201}$  201.  $\frac{1}{202}$  202.  $\frac{1}{203}$  203.  $\frac{1}{204}$  204.  $\frac{1}{205}$  205.  $\frac{1}{206}$  206.  $\frac{1}{207}$  207.  $\frac{1}{208}$  208.  $\frac{1}{209}$  209.  $\frac{1}{210}$  210.  $\frac{1}{211}$  211.  $\frac{1}{212}$  212.  $\frac{1}{213}$  213.  $\frac{1}{214}$  214.  $\frac{1}{215}$  215.  $\frac{1}{216}$  216.  $\frac{1}{217}$  217.  $\frac{1}{218}$  218.  $\frac{1}{219}$  219.  $\frac{1}{220}$  220.  $\frac{1}{221}$  221.  $\frac{1}{222}$  222.  $\frac{1}{223}$  223.  $\frac{1}{224}$  224.  $\frac{1}{225}$  225.  $\frac{1}{226}$  226.  $\frac{1}{227}$  227.  $\frac{1}{228}$  228.  $\frac{1}{229}$  229.  $\frac{1}{230}$  230.  $\frac{1}{231}$  231.  $\frac{1}{232}$  232.  $\frac{1}{233}$  233.  $\frac{1}{234}$  234.  $\frac{1}{235}$  235.  $\frac{1}{236}$  236.  $\frac{1}{237}$  237.  $\frac{1}{238}$  238.  $\frac{1}{239}$  239.  $\frac{1}{240}$  240

```
Qy 389 -----GlnIlePheLysSerAsnPheA 396
Db 2410 AAAAAAACAAACATAAAACATACATACCTAACCTTCAAACTATACATAACAACTATA 2351
Qy 396 snglyGluAsn 399
Db 2350 ATAACCAAAAC 2340

RESULT 9
US-10-240-485-79/c
; Sequence 79, Application US/10240485
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 79
; LENGTH: 6713
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-79

Alignment Scores:
Pred. No.: 0.0158 Length: 6713
Score: 111.00 Matches: 102
Percent Similarity: 33.21% Conservative: 76
Best Local Similarity: 19.03% Mismatches: 166
Query Match: 4.64% Indels: 195
DB: 6 Gaps: 21

US-09-910-186A-10 (1-450) x US-10-240-485-79 (1-6713)
Qy 4 ProPheAsnIlePheSerTyr-----ThrAsnAsnSer 14
Db 3084 CQTATTAACCTTTTCTCTACTCTTCTATCTAACCACTCTAAATTTCCCTGTAATACTCC 3025
Qy 15 LeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSer 34
Db 3024 CRT-----CCCTCCGCCATAAAT 3007
Qy 35 LeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGlu 54
Db 3006 AGCAAAACCCCAAACTATATTTCTAAACACTCTTTTCATCTCTCTTCCACTTCCC--- 2950
Qy 55 GluGlyAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGly----- 71
Db 2949 -----TCITTAATAAAATCTAAC-CCATTCTCTATATCTTTAAATACCATCTAA 2903
Qy 72 -----SerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsn 87
Db 2902 ATATAAATCATTAAACAAATCAAAACAAACGATATACTATTTCATTAATAAATACTAAA 2843
Qy 88 IleValTyrAsnSerMetTyrGluSerPheSerIleSerPheIleArgIleAsn--- 106
Db 2842 ATAACATTATATAAATTCACATCTCTTTATATAAAATCTCCACCAAAATATATATA 2783
```

```
Qy 107 -----LysTyrpValSerAsnLeuproglyTyrThrIleIle 118
Db 2782 AAAAAACATACCTCAAATAATAAAACCATATATAACAAACCTTCAATCAACATCATATA 2723
Qy 119 AspSerValLysAsnAsnSerGlyTyrpSer----- 128
Db 2722 CT-AAACACAAATAAAACCTTTCTCCAAAACCTAAACAAACAAACAAATACCC 2664
Qy 128 ----- 128
Db 2663 CTCACACACTATTATTCTAGTAAATACAAAAATCTTAACCAAACTATCAACAAAAA 2604
Qy 129 -----IleGlyIleIleSerAsn 134
Db 2603 AAAAAAATAATATCCAAATTAATAAAAAAATAATCAAAATTAATTTATTTCACAAAC 2544
Qy 135 PheLeuValPheThrLeuLysGlnAsnGluAspSerGlu-----GlnSerIle 150
Db 2543 TACATAATCTTATACCTTAAATAAATACTTAAATAATCCACCTAAACACTATAATCTAAT 2484
Qy 151 AsnPheSerTyrAspIleSerAsnAlaProGlyTyrAsnLysTyrPhePheValThr 170
Db 2483 CAATAAATTACAAATACAAATCAACATACAAATCAATTAAC---ATTATATAACA 2427
Qy 171 ValThrAsnAsnMetMetGlyAsnMetLys-----IleTyrIleAsnGlyLysLeu 187
Db 2426 ACAATAACCAATCTAATAAATAATCAAAAAACAATTCCTTTTACAAATAACACAAATA 2367
Qy 188 IleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPhe 207
Db 2366 ATACAATACCTAANA-----ATCAATTTAACTAAAAAATAAATA--- 2328
Qy 208 GluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMet 227
Db 2328 ----- 2328
Qy 228 TrpIleArgAspPheThrIlePheAlaLysGluLeu----- 239
Db 2327 -----TAICTATCAAAAAAATAAATACTTAATAAATACTTAATAAATAAATAA 2283
Qy 240 -----AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256
Db 2282 TCGAAAAAACACACAAACCCCAAAATATATATACCTCGTAATAAATACTA----- 2238
Qy 257 ValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle 276
Db 2237 -----AAAAAATAATATATTATAAATA 2214
Qy 277 AspTyrLeuAsnArgTyrMetTyrAla-AsnSerArgGlnIleValPheAsnThrArgAr 296
Db 2213 CAATAC-----TACCTAAACAAATTTACAAATTCATACAAATTCCTATCA 2169
Qy 296 GAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnTh 316
Db 2168 AAATCAAAATTAACATCTTCACGAAACAAATAAATTCCTAAATTTTATAAAACCCAC--- 2111
Qy 316 rAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLy 336
Db 2110 -----AAAAACCTTAATAACCAACAACTCTTAACAAAAAATAACAA 2067
Qy 336 sAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAs 356
Db 2066 AACTAAAAACATCATACCTACCTAACTTCAAAATTTACTA-CAAAACTATATAAATCAACAA 2008
Qy 356 pIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn----- 372
Db 2007 CAACAAATATTATACATAAAACAAACACATACCAATAAACAAATAAATAAATAATCCAC 1948
Qy 373 -----IleIlePheGlnIleGlnPrometAsnAsnThrTyrTyrTyrAlaSerGlnIlePh 391
Db 1947 ATATAAATCCATACACTTACAAACCACTAATCTATACAAACTATATAAATAAATACATAAT 1888
```

QY 391 eLysSerAsn-----PheAsnGly-----GluAsnIleSerGlyIleCysSerI 406  
||| ||| ||||| :|||: ||  
Db 1887 AAAAAAAAAAATCTCTTTAATAATAATAATACTAATAAAGAACTTA-----AT 1843  
QY 406 eGlyThrTyr-----ArgPheArgLeuGlyGlyAspTyrArgHisAsn----- 421  
||| ||| :|||: |||  
Db 1842 AACTACATACAAAAAATAAAATTAACATTTCTCACCACCTTCTCTTCCCAATTAAA 1783  
QY 422 -----TyrLeuValProThrValVallysGlnGlyAsnTyr 432  
:||||: |||  
Db 1782 TACCTTCCTCTTCTCTCTTACCTTAATAACTTAACCAAAACCTTT 1737  
RESULT 10  
US-10-240-453-25/c  
: Sequence 25, Application US/10240453  
: GENERAL INFORMATION:  
: APPLICANT: OLEK, Alexander  
: APPLICANT: PIEPENBROCK, Christian  
: APPLICANT: BERLIN, Kurt  
: TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
: TITLE OF INVENTION: Transcription  
: TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
: TITLE OF INVENTION: With DNA Transcription  
: FILE REFERENCE: 5013.1009  
: CURRENT APPLICATION NUMBER: US/10/240,453  
: CURRENT FILING DATE: 2002-10-02  
: PRIOR APPLICATION NUMBER: PCT/EP01/03973  
: PRIOR FILING DATE: 2001-04-06  
: PRIOR APPLICATION NUMBER: DE 10019058.8  
: PRIOR FILING DATE: 2000-04-06  
: PRIOR APPLICATION NUMBER: DE 10019173.8  
: PRIOR FILING DATE: 2000-04-07  
: PRIOR APPLICATION NUMBER: DE 1003529.7  
: PRIOR FILING DATE: 2000-06-30  
: PRIOR APPLICATION NUMBER: DE 10043826.1  
: PRIOR FILING DATE: 2000-09-01  
: NUMBER OF SEQ ID NOS: 350  
: SEQ ID NO 25  
: LENGTH: 5979  
: TYPE: DNA  
: ORGANISM: Artificial Sequence  
: FEATURE:  
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-453-25  
Alignment Scores:  
Pred. No.: 0.0256 Length: 5979  
Score: 108.50 Matches: 104  
Percent Similarity: 37.03% Conservative: 83  
Best Local Similarity: 20.59% Mismatches: 168  
Query Match: 4.54% Indels: 151  
DB: Gaps: 22  
US-09-910-186A-10 (1-450) x US-10-240-453-25 (1-5979)  
QY 7 IlePheSerTyrThrAsnAspSerLeuLeuLysAsp-----IleIleAsnGluTyrPhe 24  
|||: |||: |||: |||: |||: |||  
Db 4240 ATATACACCCATTTCACACAAATATTATTAAACAATAAACAATACGTAATTACAAATATCAA 4181  
QY 25 AsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValasp 44  
||||| :|||: |||  
Db 4190 AATAATATACAAACAAATTTCTCTTAATATCAAAACCATTAATAATCCACAAAACAAT 4121  
QY 45 -ThrSerGlyTyrAsnAlaGluValSerGluGlu-----Glyaspva 58  
|||: |||: |||: |||: |||: |||  
Db 4120 TACTAATTCACAAAACATATATCCTACCAACACATCCATTCACAAACACATATTCACACAC 4061  
QY 58 IGLnLeuAsnProIle---PheProPhe-----AspPheLysLeuGlySerSerG1 74  
||| ||||| ||||| ||||| |||  
Db 4060 CCAAAAAAATCCAAATCAACTTCTCTTATTCACATAAATCTACCAAAACCATCTCTCTCC 4001  
QY 74 yGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyr----- 90  
: |||: |||: |||: |||: |||: |||



GENERAL INFORMATION:  
 APPLICANT: Cech, Thomas R.  
 Lingner, Joachim  
 Nakamura, Toru  
 Chapman, Karen B.  
 Morin, Gregg B.  
 Harley, Calvin B.  
 Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
 NUMBER OF SEQUENCES: 727  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-3834

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/721,456  
 FILING DATE: 22-Nov-2000  
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/974,549A  
 FILING DATE: 19-Nov-1997

APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-Oct-1996  
 APPLICATION NUMBER: US 08/844,419  
 FILING DATE: 18-Apr-1997  
 APPLICATION NUMBER: US 08/846,017  
 FILING DATE: 25-Apr-1997  
 APPLICATION NUMBER: US 08/851,843  
 FILING DATE: 06-May-1997  
 APPLICATION NUMBER: US 08/854,050  
 FILING DATE: 09-May-1997  
 APPLICATION NUMBER: US 08/911,312  
 FILING DATE: 14-Aug-1997  
 APPLICATION NUMBER: US 08/912,951  
 FILING DATE: 14-Aug-1997  
 APPLICATION NUMBER: US 08/915,503  
 FILING DATE: 14-Aug-1997  
 APPLICATION NUMBER: WO PCT/US97/17618  
 FILING DATE: 01-Oct-1997  
 APPLICATION NUMBER: WO PCT/US97/17885  
 FILING DATE: 01-Oct-1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph Red  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-002610US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 640:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3396 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 FEATURE:

NAME/KEY:  
 LOCATION: 1..3396  
 OTHER INFORMATION: /note="HTRT-encoding sequence employing alternative codon distributions for yeast (all genes)"

SEQUENCE DESCRIPTION: SEQ ID NO: 640:  
 US-09-721-456-640

Alignment Scores:

Pred. No.:	0.0132	Length:	3396
Score:	108.00	Matches:	82
Percent Similarity:	38.28%	Conservative:	78
Best Local Similarity:	19.62%	Mismatches:	186
Query Match:	4.52%	Indels:	72
DB:	5	Gaps:	14

US-09-910-186a-10 (1-450) x US-09-721-456-640 (1-3396)

QY	13	AsnSerLeuLeuLysAspPheLeuAsnGluThrPheAsnAsnLeuAspSer-----	30
DB	3048	AAACACAGCATGAATCTATAAGCTTGCAACAAACAAATTTTATAAATATTAGTACAAC	2989
QY	31	-----LysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyr	48
DB	2988	AGTTTGCAGAAATTAACCTTGCATTCACAAACAAAGATGACA-----TTT	2941
QY	49	AsnAlaGluValSerGluGluAspValGlnLeuAsnProIlePheProPheAspPhe	68
DB	2940	CAATCTCAAAACACCAACAATTTCTTCTCATATTTTACCAGCTTTAAAAACCTCTATT	2881
QY	69	LysLeuGlySerSerGlyGluAspArgGlyLysVal-----IleValThrGluAsnGlu	86
DB	2880	AAAAGTCAAGAGCTCTAATAGAGTTCTAGCATAGAGAAATTAATCAGATTGAAC---	2824
QY	87	AsnIleValTyrAsnSerMetTyrGluSerPheSerIle-----SerPheTyr	102
DB	2823	-----TTCCAAAGTTCTAGTATCCAAACAAACACACACCATGG	2785
QY	103	IleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThr-----IleIleAspSer	120
DB	2784	AAACAAACCATGAGCTGGCATTTGACAAACAAACAGTACACCAAGCTTCTCTCAAC	2725
QY	121	ValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeu	140
DB	2724	TGGAATTAACACACAGTTTCT-----CAATTAACACACACACCATATTCTGGAAC	2671
QY	141	LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla	160
DB	2670	ACCTCTAACCAAGTTCTCAAAAAAGTTTACGATGAGTCAAAATGTGGAGTAACCAAA	2611
QY	161	ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys	180
DB	2610	AAATCATCAACCAATCTCAACACAAACCATCTCTCTTAATACCAACCAACATTTATT	2551
QY	181	IleTyrIle-----AsnGlyLysLeuIleAsp	189
DB	2550	TTCCATATCACCAATAACACAAAGACACAAACAAAGTAGACAAATAGAACCTTGTGGAAT	2491
QY	190	ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle	209
DB	2490	ACCTTGACATTGAACATAAGATTACCTCTAATTTACAGCATGATGACACATAAATCT	2431
QY	210	AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle	229
DB	2430	CAAAAA-----AACATCAACAAACACCAAG	2407
QY	230	ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
DB	2406	AGAAGCTTCATTCAAGAGAGAGATTGTTCAATACAAACAGCATCTCTCAATGGAGAGAT	2347
QY	250	SerLeuGlnTyrThrAsnValValLysAspTyrTrp-----GlyAsnAsp	264
DB	2346	TTCTTGCAATGAGCAACAAATTTGCTCATATATGTTGCAAAATCAGTCAAAAGTAGAAAC	2287
QY	265	LeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyr	284
DB	2286	ATGAGATTTTAAAGCTTTTCAACATGACCATGACGAGCTTTTGAACAAACAGCATATCT	2227
QY	285	AlaAsnSerArgGlnIleVal---PheAsnThrArgAsnAsnAsnAspPheAsnGlu	303
DB	2226	TCTAACACATAAGTATTTTGTGTTTAAATAGAGCAATAACTTCAGTCAATCTATC	2167

Qy	96	SerPheSerIleSerPheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyr	115
Db	2281	AGTTATGAA-----ATAGAGATTACAGACTCAAAAGAAATGGTTAAGGCGAAA	2328
Qy	116	ThrIleIleAsp-----SerValLysAsnAspSerGlyTrpSerIle	129
Db	2329	AAATTATTAAATTAGAACAAATTACGACAATCTTAGAAAAAACAAC	2376
Qy	130	GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer	149
Db	2377	-----TTGAAACTAGATGAAACGAACGAATTCACAAATA	2406
Qy	150	IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheVal	169
Db	2407	GCTACGATTAGCTTATAAATAGATAATAACTCTACAAATTTAATAAA	2456
Qy	170	ThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAsp	189
Db	2455	---ATRAAAAAT-----ATAATTCAGATTCTCGATCAT	2484
Qy	190	ThrLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle	209
Db	2485	GAATTAATAAGTGCAA-----	2499
Qy	210	AsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIleAsnMetTrpIle	229
Db	2499	-----	2499
Qy	230	ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
Db	2500	AAAGAAATTTCTTGCTATCAGAAATATTAGCTGGTAAA-----	2538
Qy	250	SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLys	269
Db	2539	-----AATGATTATAATTACACTA	2559
Qy	270	GluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGln	289
Db	2560	GAATAATATGTTTA--ATTATTACTTGGAAAAATAATTTTCAAGCGAATCAACGA	2616
Qy	290	IleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIle	309
Db	2617	CTTTCATTATGAGT-----GGAAATAGATATCAACTG-----	2648
Qy	310	LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe	329
Db	2650	-----ATTAGAGAGAGACCATTCTTAGGTTAAGTGGATTGAAATTGACGTGTTT	2703
Qy	330	AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla	349
Db	2704	GAT-----TTTCATTCAATTAATAATCGAGACACATA	2733
Qy	350	AspAsnHisSerThrGluAspIleTyr-----AlaIleGlyLeuArg	363
Db	2734	AGTTCATTTCAGAGAGAGAAACATTTCAAGCATCATCTAGCTTTAGCTTTAGATTAGT	2799
Qy	364	Glu 364	
Db	2794	GAA 2796	

RESULT 14  
US-10-240-485-107/c  
; Sequence 107, Application US/10240485  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; TITLE OF INVENTION: Metastasis  
; FILE REFERENCE: 5013.1007  
; CURRENT APPLICATION NUMBER: US/10/240,485  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03970

```

Db      3983  -----ATCCAACTAAATACCTTAAATAATACAACT - 3949
               |||||
QY      302  AsnGluGlyTyrLysIleIleLysArgGlyAsnThrAsnAspThrArgVal 321
               ::  |||  :::::||||:  |||
Db      3948  CAAAAAACTATACCAACTTTTAAACAATACTACCAAAAACTCAAAAAAACT - 3895
               |||  :::::||||:  |||
QY      322  ArgGlyGlyAspIleuTyPheAspMetThrIleAsnAsnLysAlaTyAsnLeuPhe 341
               ::  |||||  |||||  :::::  |||  :::
Db      3894  -----ACACTTCTATATCTACCTAAAAAACAATCTAAACACCACAAAATATCA 3844
               |||  |||||  |||||  |||||  |||  :::
QY      342  MetLysAsnGluThrMetTyAlaAspAsnHisSerThrGluAspIle - - - - -Tyr 358
               |||  |||  :::::||||:  |||
Db      3843  TCACCTTTATAATACCTCTATTTTCATCAATAACCTCTAAAAATTACCAATTAATTT 3784
               |||  |||  |||  |||  |||  |||  |||  |||
QY      359  AlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGln 378
               |||  :::::  |||  |||  |||  |||  |||  |||  |||
Db      3783  TAAATACCAATTCAAAACCCCTATTCAAANAATCAAAATAATCATCTCAATTTTCAAAC 3724
               |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      379  Pro - - - - -MetAsnAsnThrTyTyTyAlaSerGlnIlePheLys 392
               |||  ::|||  |||  |||  |||  |||  |||  :::
Db      3723  CCTAATTTTAATTTCTCAATATAATAATAAACCAATAACAATACTCATACCTATAATCC 3664
               |||  |||  |||  |||  |||  |||  |||  |||
QY      393  SerAsnPhe 395
               ::|||
Db      3663  CAACACTTT 3655
               ::|||

RESULT 15
US-10-240-454-49/c
; Sequence 49, Application US10240454
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Metabolism
; FILE REFERENCE: 5013.1010
; CURRENT APPLICATION NUMBER: US/10/240.454
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/04016
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 68
; SEQ ID NO 49
; LENGTH: 9265
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-454-49

Alignment Scores:
Pred No.: 0.116 Length: 9265
Score: 105.00 Matches: 83
Percent Similarity: 36.92% Conservative: 44
Best Local Similarity: 24.13% Mismatches: 107
Query Match: 4.39% Indels: 110
DB: 6 Gaps: 17

US-09-910-186A-10 (1-450) x US-10-240-454-49 (1-9265)

QY      133  SerAsnPheLeuValPheThrIleuLysGlnAsnGluAspSerGluGlnSerIleAsnPhe 152
               |||  |||  ::|||  |||
Db      6345  TCACCTTTTACTATATTTTAAANAATCTCTA - - - - -CTTCCCAAAAATTAATAAT 6292
               |||  |||  |||  |||  |||
QY      153  SerTyAspIleSerAsnAlaProGlyTyAsnLysTrpPhePheValThrValThr 172
               |||  |||  :::::||||  |||

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:41:56 ; Search time 43 Seconds  
(without alignments)  
1006.058 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391  
Sequence: 1 MTIPNFISYTNNSLLKDI.....NYASLLESTHGWGFPVSE 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2386	99.8	1291	2 S46431	botulinum neurotox
2	2386	99.8	1291	2 A49777	botulinum neurotox
3	1823	75.2	1285	2 S70582	botulinum neurotox
4	867.5	36.3	1276	2 S11435	botulinum neurotox
5	604.5	25.3	1297	2 S39791	neurotoxin - Clost
6	594.5	24.9	1252	2 S21178	botulinum neurotox
7	594	24.8	1251	2 JH0256	botulinum neurotox
8	593	24.8	1291	2 I40631	non-proteolytic bo
9	586.5	24.5	1296	1 BTCLAB	bontoxilysin (EC 3
10	581	24.3	1291	1 A48940	bontoxilysin (EC 3
11	569	23.8	1268	2 S34111	botulinum neurotox
12	554.5	23.2	1296	2 I40645	botulinum neurotox
13	549.5	23.0	1274	2 I40813	neurotoxin type F
14	425	17.8	1315	1 BTCLTN	tentoxilysin (EC 3
15	281.5	11.8	367	2 S48106	neurotoxin type E
16	279	11.7	366	2 S48110	neurotoxin type F
17	245.5	10.3	369	2 S48109	neurotoxin type F
18	230.5	9.6	1162	2 A47708	progenitor toxin n
19	230.5	9.6	1162	2 I40817	botulinum toxin no
20	213	8.9	1193	2 S68218	botulinum neurotox
21	199	8.3	1165	2 I40644	botulinum neurotox
22	199	8.3	1196	2 JQ1467	toxin, nontoxic co
23	199	8.3	1196	2 S46430	botulinum neurotox
24	191.5	8.0	1193	2 J4901	nontoxic-nonhemagg
25	174.5	7.3	960	2 S72284	DNA-directed RNA p
26	171.5	7.2	1844	2 D71612	hypothetical prote
27	170.5	7.1	398	1 R3BYM1	ribosomal protein
28	168.5	7.0	1817	2 H71611	probable secreted
29	168	7.0	4688	2 F62885	hypothetical prote

30	166.5	7.0	617	2 D96978	hypothetical prote
31	159	6.6	2817	2 B97033	uncharacterized pr
32	157.5	6.6	762	2 E81371	probable periplasm
33	157	6.6	1465	2 A70199	hypothetical prote
34	155.5	6.5	888	2 B82885	hypothetical prote
35	155	6.5	1711	2 T18429	hypothetical prote
36	155	6.5	3724	2 T18427	hypothetical prote
37	154.5	6.5	2013	2 C71610	probable membrane
38	153	6.4	1306	2 T28313	ORF MSV152 probabl
39	150.5	6.3	1365	2 T18419	hypothetical prote
40	150	6.3	437	2 T28180	hypothetical prote
41	148	6.2	1127	2 T28317	ORF MSV156 hypot
42	148	6.2	3973	2 B71612	hypothetical prote
43	147.5	6.2	1436	2 D71618	hypothetical prote
44	147.5	6.2	2485	1 H71621	serine/threonine-s
45	147	6.1	883	2 T18484	hypothetical prote

ALIGNMENTS

RESULT 1

S46431

botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)

N:Alternate names: BoNT/C1 protein

C:Species: Clostridium botulinum phage 1C

A:Variety: Strain C 468

C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 16-Jul-1999

R:Accession: S46431; S49107

R:Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.

Mol. Gen. Genet. 243, 631-640, 1994

A:Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxi

A:Reference number: S46426; MUID:94301293; PMID:8028579

A:Accession: S46431

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1291 <HAU>

A:Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175

A:Experimental source: strain C 468

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Superfamily: tetanus toxin

Query Match	99.8%;	Score 2386;	DB 2;	Length 1291;
Best Local Similarity	100.0%;	Pred. No. 1.7e-129;		
Matches 449;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2	TIPFNIFSYTNNSLLKDIINEYFNNDISKILSQNKNLTLDTSQYNAEVSERGDYQLN	61	
Db	843	TIPFNIFSYTNNSLLKDIINEYFNNDISKILSQNKNLTLDTSQYNAEVSERGDYQLN	902	
QY	62	PIFPDFKLGSSGDRGVIVTQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV	121	
Db	903	PIFPDFKLGSSGDRGVIVTQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV	962	
QY	122	KNSGWSIGIISNLFVFLKQNEDESEQINSYDINSNAPGYNKWFVVTNNMGNMKI	181	
Db	963	KNSGWSIGIISNLFVFLKQNEDESEQINSYDINSNAPGYNKWFVVTNNMGNMKI	1022	
QY	182	YINGKLIDITIKVKELTGINFSTIIFKIPDGLITSDSDNINMMIRDFYIFAKELDG	241	
Db	1023	YINGKLIDITIKVKELTGINFSTIIFKIPDGLITSDSDNINMMIRDFYIFAKELDG	1082	
QY	242	KDINILFNSLQTVNVKDWGNDLRYNKEYVMNIDYLNRYMYANSRQIVENTRRNND	301	
Db	1083	KDINILFNSLQTVNVKDWGNDLRYNKEYVMNIDYLNRYMYANSRQIVENTRRNND	1142	
QY	302	NEGYKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNTHSTEDI	361	
Db	1143	NEGYKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNTHSTEDI	1202	
QY	362	LRBQTKDINDNIIFOIQPMNNNTYYASOIFKSNFNGENISGICIGYFRFLGGDWTR	421	
Db	1203	LRBQTKDINDNIIFOIQPMNNNTYYASOIFKSNFNGENISGICIGYFRFLGGDWTR	1262	

QY 422 YLVPVTKOGNTASLLESTSTHWGFPVSE 450  
|||||  
Db 1263 YLVPVTKOGNTASLLESTSTHWGFPVSE 1291

RESULT 2  
A49777  
botulinum neurotoxin type C1 precursor - Clostridium botulinum phage (type C, strain c-s  
C:Species: Clostridium botulinum phage  
C:Date: 10-Mar-1994 #sequence\_revision 07-Apr-1994 #text\_change 23-Mar-2001  
C:Accession: S11291; A35396; S22166; A49777  
R:Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;  
Nucleic Acids Res. 18, 4924, 1990  
A:Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.  
A:Reference number: S11291; MUID:90370487; PMID:2204031  
A:Accession: S11291  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-84, P', 86-1291 <HAU>  
A:Cross-references: EMBL:X53751; NID:914905; PIDN:CAA37780.1; PID:g14906  
R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.  
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990  
A:Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxin  
A:Reference number: A35396; MUID:91024998; PMID:2222445  
A:Accession: A35396  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-869, R', 671-1291 <TS1>  
R:Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.  
Submitted to the EMBL Data Library, December 1991  
A:Description: Nucleotide sequence of the gene for one of the components of hemagglutinin  
A:Reference number: S22166  
A:Accession: S22166  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1291 <TS2>  
A:Cross-references: EMBL:X62389; NID:9558175; PIDN:CAA44263.1; PID:g40390  
R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.  
Appl. Environ. Microbiol. 57, 1168-1172, 1991  
A:Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and who  
A:Reference number: A49777; MUID:91282468; PMID:2059039  
A:Accession: A49777  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-607 <TS3>  
A:Cross-references: GB:D90210  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 1.7e-129;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTPFNIFSYTNSLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61  
|||||  
Db 843 TTPFNIFSYTNSLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 902  
|||||

QY 62 PIFPFDKLGSGEDRGKVIYVQENIVYNSWYSEFSISFWIRINKWVSNLPGYTIIDSV 121  
|||||  
Db 903 PIFPFDKLGSGEDRGKVIYVQENIVYNSWYSEFSISFWIRINKWVSNLPGYTIIDSV 962  
|||||

QY 122 KNSGWSIGIISNLFVTLKQNEDEQINFSYDINNAAGYNKWEFFVTNNMGNMKI 181  
|||||  
Db 963 KNSGWSIGIISNLFVTLKQNEDEQINFSYDINNAAGYNKWEFFVTNNMGNMKI 1022  
|||||

QY 182 YINGKLIDTIKVELTGINFSTKTIFFKINKIPDTGLITSDSDINMWIRDFYIFAKELDG 241  
|||||  
Db 1023 YINGKLIDTIKVELTGINFSTKTIFFKINKIPDTGLITSDSDINMWIRDFYIFAKELDG 1082  
|||||

QY 242 KDINILFNSLQYTNVVKDYGNDLYNKYEYVWINDLYNRYWYANSQIVENTRRNNDF 301  
|||||  
Db 1083 KDINILFNSLQYTNVVKDYGNDLYNKYEYVWINDLYNRYWYANSQIVENTRRNNDF 1142  
|||||

QY 302 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
|||||  
Db 1143 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202  
|||||

QY 362 LREOTKDINDMIIFQIOPMNNYYIASQIFKSNFNGENISGICSIGTYFRRLGGDWYRHN 421  
|||||  
Db 1203 LREOTKDINDMIIFQIOPMNNYYIASQIFKSNFNGENISGICSIGTYFRRLGGDWYRHN 1262  
|||||

QY 422 YLVPVTKOGNTASLLESTSTHWGFPVSE 450  
|||||  
Db 1263 YLVPVTKOGNTASLLESTSTHWGFPVSE 1291  
|||||

RESULT 3  
S70582  
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sa  
C:Species: Clostridium botulinum phage d-sa  
A:Note: host Clostridium botulinum type D (strain South Africa)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000  
C:Accession: S70582  
R:Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.  
Biochim. Biophys. Acta 1307, 123-126, 1996  
A:Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C  
A:Reference number: S70582; MUID:96283801; PMID:8679691  
A:Accession: S70582  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1285 <MOR>  
A:Cross-references: EMBL:D38442; NID:g1374775; PIDN:BA07477.1; PID:g1374776  
C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit  
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic me  
C:Superfamily: tetanus toxin  
C:Keywords: disulfide bond; neurotoxin; transmembrane protein  
F:1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>  
F:448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

Query Match 76.2%; Score 1823; DB 2; Length 1285;  
Best Local Similarity 77.7%; Pred. No. 3.8e-97;  
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

QY 2 TTPFNIFSYTNSLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61  
|||||  
Db 839 TTPFNIFSYTNSLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 898  
|||||

QY 62 PIFPFDKLGSGEDRGKVIYVQENIVYNSWYSEFSISFWIRINKWVSNLPGYTIIDSV 121  
|||||  
Db 899 PIFPFDKLGSGEDRGKVIYVQENIVYNSWYSEFSISFWIRINKWVSNLPGYTIIDSV 958  
|||||

QY 122 KNSGWSIGIISNLFVTLKQNEDEQINFSYDINNAAGYNKWEFFVTNNMGNMKI 181  
|||||  
Db 959 KNSGWSIGIISNLFVTLKQNEDEQINFSYDINNAAGYNKWEFFVTNNMGNMKI 1018  
|||||

QY 182 YINGKLIDTIKVELTGINFSTKTIFFKINKIPDTGLITSDSDINMWIRDFYIFAKELDG 241  
|||||  
Db 1019 YINGKLIDTIKVELTGINFSTKTIFFKINKIPDTGLITSDSDINMWIRDFYIFAKELDG 1078  
|||||

QY 242 KDINILFNSLQYTNVVKDYGNDLYNKYEYVWINDLYNRYWYANSQIVENTRRNNDF 301  
|||||  
Db 1079 KDINILFNSLQYTNVVKDYGNDLYNKYEYVWINDLYNRYWYANSQIVENTRRNNDF 1138  
|||||

QY 302 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 360  
|||||  
Db 1139 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 1193  
|||||

QY 361 GUREOTKD-INDNIIFQIOPMNNYYIASQIFKSNFNGENISGICSIGTYFRRLGGD-WY 418  
|||||  
Db 1194 GALDQPMDEIRKYGSIIFQPCNTFDYASQLFSSNATNRLGILSIGSYSFKLGGDYWF 1253  
|||||

QY 419 RNYLVPVTKOGNTASLLESTSTHWGFPVSE 450  
|||||  
Db 1254 RNYLVPVTKOGNTASLLESTSTHWGFPVSE 1285  
|||||

```
RESULT 4
S11455
botulinum neurotoxin type D - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S11455
R:Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegl
Nucleic Acids Res. 18, 5556, 1990
A:Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A:Reference number: S11455; MUID:91016893; PMID:2216736
A:Accession: S11455
A>Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-1276 <BIN>
A:Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 36.3%; Score 867.5; DB 2; Length 1276;
Best Local Similarity 43.2%; Pred. No. 3e-42;
Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

QY 2 TIPNIFSYTNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61
Db 839 TWPFNIFSYTNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVRGVDQLN 898
QY 62 PIFPFDKLGSGGDRGKVIYVTQENIVYNSMYESFISFWIRNKWVNLPG-YTIIDS 120
Db 899 TITVNDKFLSSGD---KIIYVNLNNILYSAIVENSVSFWIKSKDLTNSHNEYIINS 955
QY 121 VKNNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTNMAGNM 179
Db 956 IEQNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTNMAGNM 179
QY 180 KIYINGKLIDIKVKEITGINTFISKTITFEINKIPDTGLITSDSNINMWRDPIYPAKEL 239
Db 1016 KLYINGELKQSQKIEDLVKLDKTIYFVIDENTD-----ENQMLARDENIFSKEL 1067
QY 240 DGKDNILNLSQNTVVKYDNGNDLYNKYEMWYNDLYNRYMYANSRQIVFNTRRNN 299
Db 1068 SNEDINIVYEQIILRNVIKDYGNPLKEDFEYIIINDYIDRYIAPESNVLVLVQVPDRS 1127
QY 300 DFNEGYKIIIRKIRGNTDVRVRGSDILYFDMTINKAYNLFMKMETADNH---STED 356
Db 1128 KLYTGNPTITKSVDKPNYSRLNGDNIILHMLYNSRKYMIIRDITVIATOGGECQNC 1187
QY 357 IYATGLRQTKDINDNIFQIQPNMNTYYASQIFKSNFNGENISGICST-GTYRFLGG 415
Db 1188 VYALKLQSNLGNYGIG-IFSIKNIYVSNKYCSQIF-SSER-ENTMLLADIYKPWRFS--- 1241
QY 416 DWYRHNLYVPTVKQGNYSALLESTSTWGFV 446
Db 1242 --FKNAYTPVAVT--NYETKLLSTSSFWKFI 1268

RESULT 5
S39791
neurotoxin - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C:Accession: S39791
R:Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A:Title: Nucleotide sequence of the gene coding for Clostridium botulinum
A:Reference number: S39791; MUID:94052745; PMID:8268233
A:Accession: S39791
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1297 <CAM>
A:Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin
```

```
Query Match 25.3%; Score 604.5; DB 2; Length 1297;
Best Local Similarity 32.0%; Pred. No. 4e-27;
Matches 156; Conservative 87; Mismatches 172; Indels 73; Gaps 16;

QY 2 TIPNIFSYTNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61
Db 840 SIPFDLSLYTKDTLTIQVFNYYISNISSNAILSLSYRGRLIDSSGYGATMNVGSDVIEN 699
QY 62 PIFPFDKLGSGGDRGKVIYVTQENIVYNSMYESFISFWIRNKWVNLPG-YTIIDS 115
Db 900 DIGNQPKLNN--ENSNITAHQSKFVYDMSFEDNSINFVWETPKYNNNDIQTYYLQNEY 957
QY 116 TIIDSVKNNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTN 174
Db 958 TIISCIKNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTN 174
QY 175 MWGNMKIYINGKLIDITKVKELTGINFSEKTIITFEINKIPDTGLITSDSNINMWRDPIY 234
Db 1018 RLGNANTYINGSLKSEKILNLDRISSNDIDFKLNCIDDTTKF-----VWIKDFNI 1069
QY 235 FAKELDGKDNILFNSLQYTNVYKVDYNGNDLYNKYEMWYNDLYNRYMYANSRQIVFNT 294
Db 1070 FGRELNATEVSSLYWIOSNTLTKDFWGNPLRDTQYYLFPNQGMQNIYKYFSKASMET 1129
QY 295 RRNNNDNE-----GKIYIKKI---RGNTNDTRVRGSDILYFDMTINKAYNLF 341
Db 1130 APRTN-ENNAALNYQNLGLRFLFIKKASNSRNINNDNIVREGDIYILNI----- 1178
QY 342 MKNETMADNHSTEDIYAGLRQTKDINDNIIFOQPMNN--TYYYASQI---FKSNF 395
Db 1179 -----DNISDES-YRVYVLYNSKEIQQLF--LAPINDPTFYDVLQIKYIYKTY 1227
QY 396 NGENIS-----GICISIGTYRFLRGLGSDYRHH-NLYLVPTVKQGNYSALLESTS-----TH 442
Db 1228 NCQILCEKDKTKTFLGFGIKGVKDYGVVDYDNYF--CISOWYLRKRISININKLRG 1285
QY 443 WGFVPVSE 450
Db 1286 WQFIPVDE 1293

RESULT 6
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S21178; S48107; JH0257; B95294; A60027; S18111
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neuroto
A:Reference number: S21178; MUID:92174922; PMID:1541280
A:Accession: S21178
A:Molecule type: DNA
A:Residues: 1-1252 <WHE>
A:Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R:Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific
A:Reference number: S48103; MUID:94013372; PMID:8408542
A:Accession: S48107
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 616-982 <CAM>
A:Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1993
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum t
A:Reference number: JH0257; MUID:92181428; PMID:1543481
A:Accession: JH0257
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
```

Biochem. Biophys. Res. Commun. 193, 107-113, 1992

A>Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum t

A:Reference number: JH0256; MUID:92161428; PMID:1543481

A:Accession: JH0256

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-27, 'E', 29-1251 <POU>

A:Cross-references: EMBL:X62088; NID:G40379

A:Experimental source: strains ATCC 43181 and ATCC 43755

E:Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, J. Gen. Microbiol. 137, 519-525, 1991

A>Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E t

A:Reference number: S16145; MUID:91237316; PMID:2033376

A:Accession: S16145

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229, 'M', 231-252 <FUJ>

A:Cross-references: EMBL:X53180; NID:G40407; PIDN:CAA37321.1; PID:G40408

A:Experimental source: strain EL6340

C:Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter relea

C:Comment: The heavy chain mediates the binding of toxin to cell receptors while the

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

F:2-422/Product: botulinum neurotoxin type E light chain #status predicted <IG>

F:423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>

F:412-426/Disulfide bonds: #status predicted

Query Match 32.8%; Score 594; DB 2; Length 1251;

Best Local Similarity 24.2%; Pred. No. 1,6e-26;

Matches 150; Conservative

17;

Qy 2 TIPNIFSYNNSLLKDIIIEYFNFINDSKILSLQNKNTLVDTSGYNAEVSBEQVQLN 61

Db 822 SIPKLSYTDKRLIIIVFNKFKRIKSSVLLNNKDYDFTSGYDSNINNGDVKY 881

Qy 62 PIFPDEFKLGSSGDEGRKVIYQTQENIYVNSYMSFSISFWIRI----NKWVSNLPQYTI 117

Db 882 PTNKNQF--GIYDKLSEVNSIQNDYIYDNKYKNFSISFWIRIPNYDNKIVNNYEYTI 939

Qy 118 IDSVK--NKGSGIGIISNLFVTLKQKEDSQSINFSYDYSNNAPGY--NKWFEFVTVNNM 175

Db 940 INCNRDNNSGKWSLNHNHIIWTLQDSNGINKLAFNYGNANGISDIINKWIFVIINDR 999

Qy 176 MGNKIYINGKLIDTIKVKELTGINFSKITFEINKIPDTGLITSDSDNINMTIRDFYIF 235

Db 1000 LGDSKLINGNLIDKKSLINLGNTHVSDNLILFKIVNGSYTRYI-----GIYRNFIF 1050

Qy 236 AKELDGKDIINFLNSLOYTVNVKDYWGNDLRYNKYMYWNI---DYLNNRYMANSQIV 291

Db 1051 KDELDEYEQITLYNNEPNANILKDFWGNLYLLYDREYLLNVKFNNEINR--RTDSTLSI 1108

Qy 292 FNTNRN----NNDNFEGYKIIKRI--RGNTNDRTVRGGDILYFDMTINKKAYNLPKNETM 347

Db 1109 NNIRSTILLANRLYSGIKVKIQRVNNSTNDNLVRKNDQVINF--VASKYTHLL-----PL 1162

Qy 348 YADNHSTEDYIAGLREOTKDINDNIIFQIOPMNTVYVASQIFKSNFNGENISGICIG 407

Db 1163 YADYATTTNKKRTIKTSSGNRNFNOVVM-----NSVGNCTMFKNN--NGNN-----IG 1209

RESULT 8  
140631  
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum  
C:Species: Clostridium botulinum  
C:Date: 12-Aug-1996 #sequence:revision 12-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: I40631; S48103; S48104; S36015  
R:Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.  
Curr. Microbiol. 28, 101-110, 1994  
A:Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botu



A.; Residues: 634-994 <CMB>  
A.; Cross-references: EMBL:X70817; NID:9407782; PIDN:CAA50148.1; PID:9407783  
A.; Experimental source: proteolytic type B, strain NCTC 7273  
R.; Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.  
submitted to the EMBL Data Library, April 1992  
A.; Description: Partial amino acid sequence of botulinum neurotoxin type B and compared  
A.; Reference number: S21575  
A.; Accession: S21575  
A.; Molecule type: DNA  
A.; Residues: 36-217, G', 219-224, 'S', 226-246 <SZA>  
A.; Cross-references: EMBL:X11934; NID:940383; PIDN:CAA77991.1; PID:940384  
R.; Ruzano, H.; Mochida, S.; Bins, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars  
J. Biol. Chem. 267, 14721-14729, 1992  
A.; Title: Minimal essential domains specifying toxicity of the light chains of tetanus  
A.; Reference number: A42871; MUID:92340509; PMID:1634516  
A.; Accession: A42871  
A.; Status: nucleic acid sequence not shown  
A.; Molecule type: mRNA  
A.; Residues: 1-313, 'S', 315-451 <KUR>  
A.; Experimental source: strain Okra  
A.; Note: sequence extracted from NCBI backbone (NCBIP:109365)  
R.; DasGupta, B.R.; Datta, A.  
Biochimie 70, 811-817, 1988  
A.; Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with  
A.; Reference number: S07155; MUID:8900987; PMID:3139097  
A.; Accession: S07155  
A.; Molecule type: protein  
A.; Residues: 2-29, 'M', 31-45 <DAS>  
A.; Accession: S08562  
A.; Molecule type: protein  
A.; Residues: 442-463, 'R', 465-467 <DA2>  
R.; Schmidt, J.J.; Sathymoorthy, V.; DasGupta, B.R.  
Arch. Biochem. Biophys. 238, 544-548, 1985  
A.; Title: Partial amino acid sequences of botulinum neurotoxins types B and E.  
A.; Reference number: S07128; MUID:85197963; PMID:3888113  
A.; Accession: S07128  
A.; Status: preliminary  
A.; Molecule type: protein  
A.; Residues: 2-16 <SCH1>  
A.; Accession: S08573  
A.; Status: preliminary  
A.; Molecule type: protein  
A.; Residues: 2-17 <SCH2>  
A.; Accession: S08574  
A.; Status: preliminary  
A.; Molecule type: protein  
A.; Residues: 442-459 <SCH3>  
R.; Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta,  
Nature 359, 832-835, 1992  
A.; Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteo-  
A.; Reference number: S27125; MUID:93063293; PMID:1331807  
A.; Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic sy-  
C.; Genetics:  
A.; Gene: bont/b  
C.; Function:  
A.; Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2  
C.; Superfamily: tetanus toxin  
C.; Keywords: hydrolase, metalloproteinase; neurotoxin; transmembrane protein; zinc  
F; 2-441/Product: bontoxilysin B light chain #status experimental <LGHM>  
F; 442-1291/Product: bontoxilysin B heavy chain #status experimental <SVY>  
F; 230,234/Binding site: zinc (His) #status predicted  
F; 231/Active site: Glu #status predicted

Query Match 24.3%; Score 581; DB 1; Length 1291;  
Best Local Similarity 32.9%; Pred. No. 9e-26;  
Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;

QY 3 IPNINISYTNNSLLKDIINIFYPNINDSKILSLQNRKNLIVDTSGYNAEVSERGDVQLNP 62  
MPELSTYITNDTLIEFMFNKYNSEINLLNLYKDNNDILSGYGAKEVYDGVDELND 895

DB 836

QV 63 IPFFDFKLGSSEDRGRKVITQENIVNYSWYSFSTISFWIRNKW----VSNL--PGYT 116



```
Db 896 --KNQFKLTSSA--NSKIRVTQNCNIENSFLDFSVFWIRIPKYANGDIQNIHNEYT 951
QY 117 IIDSVKNSGWSIGIISNPLVFTLKQNEDESQSINFSYDISNNAPGY-NKWFVYVTVNNM 175
Db 952 IINCNNKNSGWSIGIRNRRIIWTLDINGKTKSVFFEYNIREDISEINRWFFVTINN- 1010
QY 176 MGNMKIYINGKLIDPIKVKELTGTFNFKSTIIFTEINKIPDTGLTSDSDNIN-MWIRDFYI 234
Db 1011 LNNAKIYINGKLESWTDIKDIREVIANGELIFKLD-----GDIDRQPIWAKVFSI 1061
QY 235 FAKELDGKDINILFNSLOYTVNVKDYWGNDLRYNKEYMYMVIDLYNRYMYANS-----RQI 290
Db 1062 FNTLSQSNIERYKIQSYSEYKLDWFCNGLPMYNKEYMFNAGNKNSYIKLKKDSPVGEI 1121
QY 291 VFNTRNNN-----DFNEGYKIIKIRGNT--NDTRVRGGDILYFD-MTINKN--- 336
Db 1122 LTRSYNQNSKYINVRDYIGEKFIIR-KNSQSINDDIVRKEDYIYLDFFNLQWVRV 1180
QY 337 -AYNLFMKNETMYADNHSTEDIYAIGLEQTKDINDNIIFOIQPMNNITYYASQ-IFKSN 394
Db 1181 YTYKYFKKEE-----EKLFLAPISDSDEFYN--TIQKEYDEQPTYSCQLLFKK- 1227
QY 395 FNGENISGICSTGYRFRLLG-----DWYRHNYLVPTVKOGNYASLLESTST 441
Db 1228 -DEESTDELGLIGHRFYESGIVPEEYKDYFCISKWY-----LKEVRKPYNLKL---GC 1278
QY 442 HWGFVPVSE 450
Db 1279 NWQFIPKDE 1287

RESULT 11
botulinum neurotoxin type F - Clostridium baratii
C:Species: Clostridium baratii
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
R:Accession: S33411; S31860
R:Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.;
FEMS Microbiol. Lett. 108, 175-182, 1993
A:Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin
A:Reference number: S33411; MUID:9325228; PMID:8486245
A:Accession: S33411
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1268 <THO>
A:Cross-references: EMBL:X68262; NID:949138; PIDN:CAA48329.1; PID:949139
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 23.8%; Score 569; DB 2; Length 1268;
Best Local Similarity 32.7%; Pred. No. 4.3e-25;
Matches 154; Conservative 82; Mismatches 167; Indels 68; Gaps 17;

QY 2 TPFNFTSTNLSLKDIINEYFNNDKILSQNRKNTLVDSGYNAEYSEGDVQLN 61
Db 832 SIFFELSEYTNPKDILHILRFYKRIIDSSILNKNKYENRFDSSGYSNISINGDIYII 891
QY 62 PIPFPFKLGGSGEDRGKVIYTONENIYVNSMYBSFSISFWIRINKW--VSNLPG-YTII 118
Db 892 STNRNCPGYISS--RLSEVNTQNTNIYNSYQNFVSFWVRIPKYNLKNLNNETII 949
QY 119 DSVK-NNSGWSIGITISNPLVFTLKQNEDESQSINFSYDISNNAPGYNKWFVYVTVN 173
Db 950 NCMRRNNSGWSIGISLNNIITWLTQTNQKLVFN--QIMDISY---INKWTFVITN 1006
QY 174 NMGNMKIYINGKLIDPIKVKELTGTFNFKSTIIFTEINKIPDTGLTSDSDNINMWIRDFY 233
Db 1007 NRUGHSKLYINGNLIDQKLSILNGLNIHVDDNLFIKIVGCDNTRYV-----GIRYFK 1057
QY 234 IFAKELDGKDINILFNSLOYTVNVKDYWGNDLRYNKEYMYMVIDLYNRYMYANSRQVFN 293
Db 1058 IFNMELDKTEIETLYHSEDPSTILKDFWGNVLYNKKYIYLLNLKPNMSVTKNSDILNIN 1117
```

```
QY 294 TRP-----NNDFNEGYKIIKIRGNTN---DTRVRGGDILYFDMTINNKAYNL 340
Db 1118 RORGYSKTNIFSNARLYTGVVEIIRKV-GSTDTSTNDFNVRKNDTVIINVDGNSBIQL 1176
QY 341 FMKNETMYADNHSTEDIYAIGLE-OTKRDINDNIIFOIQPMNNITYYASQIFKSNFNGEN 399
Db 1177 -----YADYSTAVENTIKLRISNSYNSN--QMIIMDSIGDNCNTMNFKN-NGND 1225
QY 400 ISGICSTGYRFRLL-----GGDWYRHNYLVPTVKOGNYASLLESTSTHWGFV 446
Db 1226 -----IGLGFHLNVLVASSWYIKYKIRNTRNNGCF-----WSFI 1260

RESULT 12
botulinum neurotoxin type A - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
R:Accession: I40645
R:Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type
A:Reference number: I40645; MUID:94143603; PMID:8310180
A:Accession: I40645
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1296 <RES>
A:Cross-references: EMBL:X73423; NID:9507070; PIDN:CAA51824.1; PID:9507071
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 23.28; Score 554.5; DB 2; Length 1296;
Best Local Similarity 30.8%; Pred. No. 3e-24;
Matches 147; Conservative 81; Mismatches 183; Indels 67; Gaps 12;

QY 3 IPFNFTSTNLSLKDIINEYFNNDKILSQNRKNTLVDSGYNAEYSEGDVQLNP 62
Db 849 IPFQLSKYVDNKKLLSTETEKIKVNTSILSIYKKDLDLDSRYGAKINIGRVIYDS 908
QY 63 IPFPFKLGGSGEDRGKVIYTONENIYVNSMYBSFSISFWIRINKWVSNL---PGYTIID 119
Db 909 IDKNQIKLINLESSTIEVIL--KNAIYVNSMYENFSTFWIKIPKFSKINLANNEYTIIN 966
QY 120 SVKNSGWSIGITISNPLVFTLKQNEDESQSINFSYDISNNAPGY-NKWFVYVTVNNMGN 178
Db 967 CIENSGHKVSLNYGEIITWLTQDNKQNIQVVKYSQMVNLSDYINRIFVITNNRLTK 1026
QY 179 MKIYINGKLIDPIKVKELTGTFNFKSTIIFTEINKIPDTGLTSDSDNINMWIRDFYIPAKE 238
Db 1027 SKIYINGRLIDQKPISNLGN-HASNKIMFKLDGCRD-----PRYIMIKYFNLPDKE 1078
QY 239 LDGKDINILFNSLOYTVNVKDYWGNDLRYNKEYMYMVIDLYNRYMYANS----- 287
Db 1079 LNEKEIKLDYSQSGSILKDFWGNLYQDPYVYMLNLFDPENKYVDYNNIGIRGYMYLKG 1138
QY 288 --RQIVFNTNRNNDFNEGYKIIKIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNE 345
Db 1139 PRGSVVTNIYLNSTLYEGTKFIKKVASGNEEDIVRNDRVYINVVYKNEYRL----- 1193
QY 346 TWYADNHSTEDIYAIGLEQTKDINDNIIFOI-----QPMNNTIYASQIFKSNFNG 397
Db 1194 ----ATNASQAGVEKILSALEIPDVGN--LSQVYVVKSKDDQGIKRNK-----CKMNLQD 1241
QY 398 ENISGTCSTGYRFR-----RLGGDWYRHNYLVPTVKOGNYASLLESTSTHWGFVPSVSE 450
Db 1242 NNGNDIGFIFGLHYDNIKLVASNWN-----RQVKGAS--RTFGCSWEFIPVDD 1289
```

```
RESULT 13
I40813
botulinum neurotoxin type F - Clostridium botulinum
C:Species: Clostridium botulinum
```



C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999  
C/Accession: 140813; S48108  
R/Author: A. K.; Richardson, P. T.; Allaway, D.; Collins, M. D.; Roberts, T. A.; Thompson, D. E.  
R/Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.  
A/Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.  
A/Reference number: 140844  
A/Accession: 140813  
A/Status: preliminary; translated from GB/EMBL/DDDB  
A/Molecule type: DNA  
A/Residues: 1-1274 <RES>  
A/Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867  
R/Campbell, K. D.; Collins, M. D.; East, A. K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A/Title: Gene probes for identification of the botulinum neurotoxin gene and specific isoforms.  
A/Reference number: S48103; MUID:94013372; PMID:8408542  
A/Accession: S48108  
A/Status: preliminary; translation not shown  
A/Molecule type: DNA  
A/Residues: 634-1002 <CAM>  
A/Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789  
C/Superfamily: tetanus toxin  
C/Keywords: neurotoxin

Query Match 23.08; Score 549.5; DB 2; Length 1274;  
Best Local Similarity 31.68; Pred. No. 5.7e-24;  
Matches 147; Conservative 83; Mismatches 180; Indels 55; Gaps 15;

QY 2 TIFNIFSYNNLSKLDINIEYFNNDISKILSONKNTLVDTSGYNAEVEGDVQLN 61  
DB 840 SIPFELSYNTDKILLIYFNRLYKIKDKSSILDMRYENKFDIDTSGYNSINNGVYII 899  
QY 62 PIFPDKLGSSGDEGRGVIVTONENIYVNSMYSFISFWIRNKW--VSNLPGVTII 118  
DB 900 STNRQF--GIYNSLSEVIAQNDIIYNSRYQNFISFWIRPKYKPMHNRFTII 957  
QY 119 DSV-KNNSGWSIG--IISNLFVLTQKNESEISNFSYDISNAPGY-KNMFVFTVN 173  
DB 958 NCMGNNNSGWSIKSLATVRDCEIITLQDTSGNKENLFRVEELNRIYNNKFTVITN 1017  
QY 174 NMGNKVIYNGKLDITKVRELGTGINSKTIPTKINKIPTDGLTSDSNINMWRDFY 233  
DB 1018 NRIISRYINGNLIVKESISNLGDIHVSMDILFK-----IVGCDTETVYGIIRFK 1068  
QY 234 IFAKELGDKDINILFNSLYQYTNVVDYNGDNLRYNKEYYMNIDILRYNRYANSRQIVFN 293  
DB 1069 VFNTLDTKEIETLYSNPDPSILKNYGNVLLYNNKKYLLNLRKDKYITLNSGILNIN 1128  
QY 294 TRRN-----NDFENEGYKIIIR--IRGNTDVRGCDILYDMDINNKAYNLFM 342  
DB 1129 QQRGVTEGSVFLNYKLYGEVYIIRKNGPIDISNTDFVRKNDLAYINVDVGRVYRL-- 1186  
QY 343 KNETYADNRHSTEDIVAGLRQTRKDINDNIIFQIPQMNNFYVYASQIFKSNFNGENISG 402  
DB 1187 ----YADTKSEKII-----RTSNLNDLS-LGQIIVMDISGNCTMNFQNN-NGSN-- 1231  
QY 403 ICSIGYRFR---LGGDWYRNLYVPVQKGNVSLLESSTHW 443  
DB 1232 ---IGLLGFHNNLVASSWYNNIRNRTSSNGCFWSSI-SKENGW 1272

RESULT 14  
BTCLTN  
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani  
N/Alternate names: tetanus neurotoxin  
C/Species: Clostridium tetani  
C/Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 03-Jun-2002  
C/Accession: A25689; A25757; A25194; A60759; S69348; S09364  
R/Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.  
EMBO J. 5, 2495-2502, 1986  
A/Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxin.  
A/Reference number: A25689; MUID:87053614; PMID:3536478  
A/Accession: A25689  
A/Molecule type: DNA

A/Residues: 1-1315 <EIS>  
A/Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770  
R/Fairweather, N. F.; Lyness, V. A.  
Nucleic Acids Res. 14, 7809-7812, 1986  
A/Title: The complete nucleotide sequence of tetanus toxin.  
A/Reference number: A25757; MUID:87040747; PMID:3774547  
A/Accession: A25757  
A/Molecule type: DNA  
A/Residues: 1-1315 <FAI>  
A/Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774  
A/Experimental source: strain CN3911  
R/Fairweather, N. F.; Lyness, V. A.; Pickard, D. J.; Allen, G.; Thomson, R. O.  
J. Bacteriol. 165, 21-27, 1986  
A/Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C 1  
A/Reference number: A25194; MUID:86085672; PMID:3510187  
A/Accession: A25194  
A/Molecule type: DNA  
A/Residues: 743-1315 <FA2>  
A/Cross-references: GB:ML2739; NID:g144920; PIDN:AAA23282.1; PID:g144921  
A/Accession: B25194  
A/Molecule type: protein  
A/Residues: 865-894 <FA3>  
R/Matsuda, M.; Lei, D. L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.  
Infect. Immun. 57, 3588-3593, 1989  
A/Title: Isolation, purification, and characterization of fragment B, the NH-2-terminus of tetanus toxin.  
A/Reference number: A60759; MUID:90035436; PMID:2478476  
A/Accession: A60759  
A/Molecule type: protein  
A/Residues: 461-475 <MA>  
R/Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.  
J. Immunol. 142, 394-402, 1989  
A/Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.  
A/Reference number: JS0098; MUID:89093918; PMID:2463305  
A/Contents: annotation; epitope region  
R/Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P. P.; DasGupta, R.  
Nature 359, 832-835, 1992  
A/Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of syntaxin.  
A/Reference number: S27125; MUID:93063293; PMID:1331807  
A/Contents: annotation  
R/de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.  
Eur. J. Biochem. 229, 61-69, 1995  
A/Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.  
A/Reference number: S69348; MUID:95262688; PMID:7744050  
A/Accession: S69348  
A/Molecule type: protein  
A/Residues: 2-31 <DEF>  
C/Comment: The source of this protein was an extrachromosomal plasmid.  
C/Comment: The precursor is cleaved by endogenous proteinase activity to form light (L) and heavy (H) chains. The amino end of the heavy chain (fragment B) binds to ganglioside GM1. The amino end of the light chain (fragment C) binds to ganglioside GM1. The amino end of the light chain (fragment C) binds to ganglioside GM1. The amino end of the light chain (fragment C) binds to ganglioside GM1.  
C/Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized, and inhibits neurotransmitter release by proteolytic cleavage of presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of presynaptic neurons.  
C/Function:  
C/Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin.  
C/Superfamily: tetanus toxin  
C/Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
F:2-457/Product: tetoxylysin light chain (fragment A) #status predicted <TTL>  
F:461-1315/Product: tetoxylysin heavy chain (fragment B.C) #status experimental <TTH>  
F:461-864/Domain: channel forming (fragment B) #status predicted <XHB>  
F:865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>  
F:233/237/Binding site: zinc (His) #status predicted  
F:234/Active site: Glu #status predicted

Query Match 17.88; Score 425; DB 1; Length 1315;  
Best Local Similarity 26.48; Pred. No. 8.5e-17;  
Matches 134; Conservative 90; Mismatches 169; Indels 114; Gaps 21;

QY 3 IPF----NIFSTNNLLKLDIINEYFNNDISKILSONKNTLVDTSGYNAEVEGDV 58  
DB 859 IPFSYKSLDCVNDDEIDVI-----LKKSTLLINDNDIISDISGFSNVITYPDA 912  
QY 59 QLNPIFPDFKLGSGE-----DRGKVIQENIYVNSMYSFISFWIRNK-WVS 110

Search completed: November 7, 2002, 14:46:49  
Job time : 53 secs



CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
CC -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C  
CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE  
CC BACTERIOPHAGE.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X66433; CAA47060.1; -  
CC EMBL; X73793; CAA51313.1; -  
CC DR EMBL; X53751; CAA37780.1; -  
CC DR EMBL; D90210; BAA14235.1; -  
CC DR EMBL; X62389; CAA44263.1; -  
CC PIR; S11291; S11291.  
CC PIR; A35396; A35396.  
CC PIR; A43503; A43503.  
CC DR HSSP; P10845; 3BTA.  
CC DR MEROPS; M27.002; -  
CC DR InterPro; IPR000395; Bontoxilysin.  
CC DR InterPro; IPR000130; Zn\_Wipeptidase.  
CC DR Pfam; PF01742; Peptidase\_M27; 1.  
CC DR PRINTS; PR00760; BONTOXILYSIN.  
CC DR PRODOM; PD001963; Bontoxilysin; 1.  
CC DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
FT INIT\_MET 0 0  
FT CHAIN 1 448 BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.  
FT CHAIN 449 1290 BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.  
FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DISULFID 436 452 INTERCHAIN (PROBABLE).  
FT CONFLICT 84 84 P -> T (IN REF. 2).  
SQ SEQUENCE 1290 AA; 148734 MW; 71FB6379F9712958 CRC64;  
Query Match 99.8%; Score 2386; DB 1; Length 1290;  
Best Local Similarity 100.0%; Pred. No. 8.2e-135;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
2 TIFPFIYSYNNLLKDTINEYFNINDSKILSLQNKNTLVDTSYNAEVSSEGDVQLN 61  
842 TIFPFIYSYNNLLKDTINEYFNINDSKILSLQNKNTLVDTSYNAEVSSEGDVQLN 901  
62 TIFPFDKLGSGEDRGKVIYQNTENIVYNSMYESFSISFIRINKWNSLPGYTIIDSV 121  
902 TIFPFDKLGSGEDRGKVIYQNTENIVYNSMYESFSISFIRINKWNSLPGYTIIDSV 361  
122 KNNCSWSGIISNLFVFLTKONESEOSINFSYDINSNAPGYNKWFVTVYNNMGNMKI 181  
962 KNNCSWSGIISNLFVFLTKONESEOSINFSYDINSNAPGYNKWFVTVYNNMGNMKI 1021  
182 YNGKLIDTIVKELTGINFSKTITFEINKIPDGLTSDSDINMWIRDFYFAKELDG 241  
1022 YNGKLIDTIVKELTGINFSKTITFEINKIPDGLTSDSDINMWIRDFYFAKELDG 1081  
242 KDINILFNSQYTNVVDYWGNDLRYNKEYYWNIDYLNRYMTANSRQIVFNRNNDF 301  
1082 KDINILFNSQYTNVVDYWGNDLRYNKEYYWNIDYLNRYMTANSRQIVFNRNNDF 1141  
302 NQGYKIIIRKRGNTNDRVRGGDILYFDMTINKKAYNLFMKNETMYADNHSTEDYVIG 361  
1142 NQGYKIIIRKRGNTNDRVRGGDILYFDMTINKKAYNLFMKNETMYADNHSTEDYVIG 1201

QY 362 LRQTKDINDNIIQIOPMNTYYIASQIFKSNFNGENISGICSTGYRFRGLGGDWYRHN 421  
Db 1202 LRQTKDINDNIIQIOPMNTYYIASQIFKSNFNGENISGICSTGYRFRGLGGDWYRHN 1261  
QY 422 YLPTVKQGNYSALLESTSTHMGFVPSV 450  
Db 1262 YLPTVKQGNYSALLESTSTHMGFVPSV 1290  
RESULT 2  
BXD\_CLOBO  
ID BXD\_CLOBO STANDARD; PRT; 1276 AA.  
AC PI9321;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BONT/D)  
DE Botulinum neurotoxin type D.  
GN BOTD.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBL\_TaxID=1491;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BVD/-3;  
RX MEDLINE=91016853; PubMed=2216736;  
RA Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,  
RA Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;  
RA "Nucleotide sequence of the gene encoding Clostridium botulinum  
RA neurotoxin type D";  
RL Nucleic Acids Res. 18:5556-5556(1990).  
RN [2]  
RC SEQUENCE FROM N.A.  
RP STRAIN-CB16;  
RX MEDLINE=93042272; PubMed=1420572;  
RA Sunagawa H., Ohyanagi T., Watanabe T., Inoue K.;  
RA "The complete amino acid sequence of the Clostridium botulinum type D  
RA neurotoxin, deduced by nucleotide sequence analysis of the encoding  
RA phage d-16 phi genome";  
RN J. Vet. Med. Sci. 54:905-913(1992).  
RN [3]  
RC PARTIAL SEQUENCE.  
RP STRAIN-D-SA, and D-1873;  
RX MEDLINE=89339741; PubMed=2668193;  
RA Morishita K., Syuto B., Kubo S., Oguma K.;  
RA "Molecular diversity of neurotoxins from Clostridium botulinum type D  
RA strains";  
RL Infect. Immun. 57:2885-2891(1989).  
RN [4]  
RC IDENTIFICATION OF SUBSTRATE.  
RX MEDLINE=94230352; PubMed=8175689;  
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,  
RA Roques B., Fyke E.M., Suedhof T.C., Jahn R., Niemann H.;  
RA "Cleavage of members of the synaptobrevin/VAMP family by types D and  
RA F botulinum neurotoxins and tetanus toxin";  
RL J. Biol. Chem. 269:12764-12772(1994).  
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF  
CC SYNAPTOSOMALIN-1 AND -2.  
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
CC detected action on small molecule substrates.  
CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
CC -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D  
CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE  
CC BACTERIOPHAGE.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X54254; CAA38175.1; -  
CC EMBL: S49407; AAB2444.1; -  
CC PIR: S11455; S11455.  
CC HSP: P10845; 3BTA.  
CC  
CC InterPro: IPR000395; Bontoxilysin.  
CC InterPro: IPR000130; Zn\_Mtpeptidse.  
CC Pfam: PF01742; Peptidase\_M27; 1.  
CC PRINTS: PR00760; BONTOXILYSIN.  
CC ProDom: PD001963; Bontoxilysin; 1.  
CC PROSITE: PS00142; ZINC\_PROTEASE; 1.  
CC Botulinum neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
CC CHAIN 1 442 BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.  
CC METAL 229 230 ZINC (CATALYTIC) (BY SIMILARITY).  
CC ACT\_SITE 230 233 ZINC (CATALYTIC) (BY SIMILARITY).  
CC METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
CC D-SULFID 437 450 INTERCHAIN (PROBABLY).  
CC VARIANT 15 16 ND -> PV (IN STRAIN D-SA).  
CC VARIANT 17 18 ND -> LQ (IN STRAIN D-1873).  
CC VARIANT 452 452 K -> Q (IN STRAIN D-SA).  
CC VARIANT 457 457 R -> T (IN STRAIN D-SA).  
CC VARIANT 457 457 R -> F (IN STRAIN D-1873).  
CC VARIANT 462 462 A -> D (IN STRAIN D-1873).  
CC VARIANT 489 489 K -> N (IN STRAIN C316).  
CC VARIANT 544 544 N -> K (IN STRAIN C316).  
CC VARIANT 1122 1122 Q -> R (IN STRAIN C316).  
CC SEQUENCE 1276 AA; 146871 MW; C1EC50F46C8233E2 CRC64;  
  

Query Match 36.3%; Score 867.5; DB 1; Length 1276;  
Best Local Similarity 43.2%; Pred. No. 2e-44;  
Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;  
  

QY 2 TIFPFIYSYTNNSLLKDIINEYFNNDKSLQKRNKTLVDTSGYNAEYSEGDYQLN 61  
DB 839 TMPFNIFSYTNNSLLKDIINEYFNNDKSLQKRNKTLVDTSGYNAEYSEGDYQLN 898  
QY 62 PIPFDFKLGSGEDRGKVIYTONENIVYNSMYESFSFWIRLNKWSNLPGLPG-YTIIDS 120  
DB 899 TITNDFKLGSSGD---KIIVNLLNLSAIYENSVSFWIKISKDUTNSHNHYTIINS 955  
QY 121 VKNSGWSGIISNLFVTLKQNEDESEQINFSDISNNAPGY-NKPFVTVTNMAGNM 179  
DB 956 IEQNSGKLCIRNGNTEWILQDVNRKYKSLIFDYSELSHTGYTNKFEVITINMGYM 1015  
QY 180 KYINGKLDTIKVKELTGINSKIITPEINPIPTGLTSDSNINWIRDFYFAKEL 239  
DB 1016 KLYINGELQSQKIEDLVKDKTKTTFVGDENID-----ENQMLWIRDFNIFSKEL 1067  
QY 240 DGKDINTLFSLOYTNVWKDYNGNDRYKEXYMYNIDVLYNRYMANSRQIVFVTRNNN 299  
DB 1068 SNEDINIVYEGQLRNVIKDYNGNPLKFTTEYIINDVINDVIDYAPESNVLYLVQYDPS 1127  
QY 300 DFENGYKIIIRKTRGNWNTDVRVGGDILYFDMTINKKAYNFKMKETWYADNH---STED 356  
DB 1128 KLYATGNPITIKSYSDKNPYSRILNGNIIHLNLYNSRKYMIIRDYTDYTAQGGECSONC 1187  
QY 357 IYALGLEQKQKINDNIIIFIQIPMNNITYYASQIFKSNFNGENISSICSI-GTYFRRLGG 415

DB 1188 VYALKQSLNGLGIG-IFSINKVSNKVCQIF-SSFR-ENTMILLADYKWRFS--- 1241  
QY 416 DWYRNYLVPTVKQGNVYASLLESTSTHGFV 446  
DB 1242 --FKNAYTPVAVT--NYETKLLSTSSFWKFI 1268  
  

RESULT 3  
EXG\_CLOBO STANDARD; PRT; 1296 AA.  
AC Q60393;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (Bont/G)  
DE (Bontoxilysin G).  
DE BORG.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=113 / 30;  
RX MEDLINE=94092745; PubMed=8268233;  
RA Campbell K., Collins M.D., East A.K.;  
RT "Nucleotide sequence of the gene coding for Clostridium botulinum  
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison  
RT with other clostridial neurotoxins";  
RL Biochim. Biophys. Acta 1216:487-491(1993).  
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
CC ENDOPEPTIDASE.  
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
CC detected action on small molecule substrates.  
CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X74162; CAA52275.1; -  
CC HSP: P10845; 3BTA.  
CC MEROPS: M27.002; -  
CC InterPro: IPR000395; Bontoxilysin.  
CC InterPro: IPR000130; Zn\_Mtpeptidse.  
CC Pfam: PF01742; Peptidase\_M27; 1.  
CC PRINTS: PR00760; BONTOXILYSIN.  
CC ProDom: PD001963; Bontoxilysin; 1.  
CC PROSITE: PS00142; ZINC\_PROTEASE; 1.  
CC Neurotoxin; Hydrolase; Metalloprotease; Zinc.  
CC INIT\_MET 0 0 BY SIMILARITY.  
CC CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.  
CC METAL 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.  
CC ACT\_SITE 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
CC ACT\_SITE 230 230 BY SIMILARITY.  
CC METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).

FT	DISULFID	435	449	INTERCHAIN (PROBABLE).
SQ	SEQUENCE	1296 AA;	149013 MW;	DC8E47E15F665C31 CRC64;
	Query Match	25.3%;	Score 604.5;	DB 1; Length 1296;
	Best Local Similarity	32.0%;	Pred. No. 9.1e-29;	
	Matches 156;	Conservative 87;	Mismatches 172;	Indels 73; Gaps
QY	2	TIPFNFSYNLSLKDIINEYFNINDKILSLQNQRKNTLVDTSGVNAEVSEBGDVLN	61	
DB	839	SIFPDSLTKTILLIQVNNNISISSNALISLSYRGRLLIDSSGATNMNVGSIFN	898	
QY	62	PIPFDFKLGGSGEDRGKIVITQENIYVNSMYESFSFWIRINKWVSN----	LPGY 115	
DB	899	DICNGQFKLNS--ENSNTTAHOSKFVVYDSMFDNFSNEWVRGPKYNNNDIQTYLQNEY	956	
QY	116	TIDSVKNSGWSIGIISFLPYTLKQNEDESQSINFSDYSINNAPGY-KNKFPVTVTNN	174	
DB	957	TIISCLKNSGMKVSIGKRRIITWLIDVNAKSISFFSYISKDLNSDYINKWFSTITND	1016	
QY	175	MMGNMKIYINGKLDIRKYELGTINFSKTTFEINKPDTGLTISDSDNIMWRIFYI	234	
DB	1017	RLGNANIYINGSLKKSEKILNLDRINSNDIDFKLINCTDTKE-----	VWKDFNI 1068	
QY	235	FAKELDGKDINILFNSLOTVNVKQYWGNDLYRKKEYVMVIDYLNRMYANSQIVENT	294	
DB	1069	FGRELAATEVSSLWTQSSTNLTKDPFGNPURDYQYYLFNQGMQNTIYIKFYKSASGET	1128	
QY	295	RRRNNDPNE-----GYKIITIKRI---RGNTNDRVRGGDILFDMTINNKAYNLF	341	
DB	1129	APRTN-FNNAAINVQNYLGLRFILKKASNSRNINDNIVREGDYILNI-----	1177	
QY	342	MKNETMVADNHSTEDIYAIGLBQKQTDNDNIIFOIQPMNN--TYVASQI----	FKSNF 395	
DB	1178	-----DNISDES-YRYIVLVNSKEIQQLF--LAPINDPTFYDVLQIKKYEITY	1226	
QY	396	NGBNIS-----GICSIGTYFRLGGDWYRE-NVLVTVKQGNVASFLESTS-----	TH 442	
DB	1227	NQQLICEKDTKTGLGICGCKGVKDYGYWDYTDNYF--CISQWYLRISENINKLRGCN	1284	
QY	443	WGEVPVSE	450	
DB	1285	WQFIPTVE	1292	

```

RESULT 4
BXE_CLOBU
ID   BXE_CLOBU          STANDARD;          PRT;   1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DS   (Bontoxilsin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinum neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).

```

[2]  
RN SEQUENCE OF 1-251 FROM N.A.  
RP  
RC STRAIN=BL6340;  
RX MEDLINE=91237316; PubMed=2033376;  
FY Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,  
Yokosawa N., Yashiki T., Oguma K.;  
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum

type E toxin gene from Clostridium butyricum strain Bf6340." ;  
J]. Gen. Microbiol. 137:519-525(1991).  
[3]  
SEQUENCE OF 1-48.  
RC STRAIN=5262;  
RA "Gimenez J., Foley J., Dasgupta B.R.;  
RA "Neurotoxin type E from Clostridium botulinum and C. butyricum;  
RC partial sequence and comparison.";  
RL FASEB J. 2:A1750-A1750(1988).  
RT  
CC -] FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
CC ENDOPEPTIDASE.  
CC  
CC -] CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
CC neuroexcitatory apparatus, synaptobrevins, SNAP25 or syntaxin. No  
CC detected action on small molecule substrates.  
CC  
CC -] SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC  
CC -] SUBCELLULAR LOCATION: Secreted.  
CC  
CC -] MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
CC  
CC -] SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X52088; CA443998.1; -;  
DR EMBL; X53180; CA437321.1; -;  
DR PIR; JH0256; JH0256.  
DR PIR; S16145; S16145.  
DR HSPSP; P10845; 3BTA.  
DR MEROPS; M27.002; -;  
DR InterPro; IPR000395; Bontoxilysin.  
DR InterPro; IPR000130; Zn\_Mtpeptidase.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR PRINTS; PR00760; BONTOXILYSIN.  
DR ProDom; PD001963; Bontoxilysin; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
KW INIT\_MET 0  
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.  
FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.  
FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 212 212 BY SIMILARITY.  
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DISULFID 411 425 INTERCHAIN (PROBABLE).  
FT CONFLICT 229 229 K -> M (IN REF. 2).  
SQ SEQUENCE 1250 AA; 143265 MW; 8171B5B2C2312857 CRC64;

Query Match 24.8%; Score 594; DB 1; Length 1250;  
Best Local Similarity 32.2%; Pred. NO. 3.7e-28;  
Matches 150; Conservative 90; Mismatches 160; Indels 66; Gaps 17;

Qy	2	TIPNIFSYNNLSLLKDIINEYFNNINDSKILSLQNKKTLYDTSGYNAEVEBEGDQVLN	61
Db	821	SIPEKLSYDDKLLSYFNKFFRKSSSVLNRKNDYDTSGISNININGDYKY	880
Qy	62	PIEPFDEKLGSSGDRKVTVTQENIVNNSMYESTPSFIWRI-----NKWYSNLPGYTI	117
Db	881	PTNKNQF--GIYNDKLSSEVNISSDIYIDNKYKNFSIFSWFRIPNIDYDNKIVYNNXETI	938
Qy	118	IDSVK--NNSGWS--GISLNFVLVFTLKQNEDESEQINFSYISNNAPGY--NKWFEVTVNNK	175
Db	939	INCNRDNNSGWYSLNHNETIWTLDNSGINOKLAFNYGNANCISDYINKFLVFTITNDR	998

QY 176 MGNKIYINGKLDTIKVRELGTGINSFKTITEINKIPDTGLTSDSDNNINWIRDFYIF 235  
 Db 999 LGSOKIYINGNLDDKSIINLGNHIVSDNLEFKIYVNCSTYRI-----GIRYFNIF 1049  
 QY 236 AKELGDKDINILFNSLOYTNYVKDYNGNDRNKEYYMWNI-----DYNLRWYANSROIV 291  
 Db 1050 DKELDETEIOTLYNNEPNANILKDFWGNLYLDKEYILLNVLKPNFNIR--RTDSTLSI 1107  
 QY 292 FNTFRN---NNDNEGYKIIKRI-RGNTNDRVRGDDILYFDMTINKKAYNLKMKNETM 347  
 Db 1108 NNIRSTILLANRLYSGIKVKIQWNSSTNDMLVRKNDQVYINF-VASKTHLL-----PL 1161  
 QY 348 YADNHSTEDIYAGLRBQRTKQINDNTIIFOQPMNNTYYVYASOIFKSNFNGENISICSIG 407  
 Db 1162 YADTATNKTEKTIKISSGNGRNFQVVM-----NSVGNCTMNFKN--NGNN-----IG 1208  
 QY 408 TYFR-----LGGQWYRHNLYVPVKOGNYASLLEPSTH---WGFW 446  
 Db 1209 LLGPKADTVVASTWY-----YTHMRDNTNSNGFFWFI 1241

RESULT 5  
 BXAL\_CLOBO STANDARD; PRT; 1295 AA.  
 AC P10845; P18639; P01561;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)  
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-  
 chain; Botulinum neurotoxin A, heavy-chain].  
 GN BOTA OR BNA OR ATX.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NTC 2916;  
 RX MEDLINE=90235864; PubMed=2185020;  
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,  
 RA Shone C.C., Atkinson T., Melling J., Minton N.P.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type A  
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding  
 RT gene";  
 RL Eur. J. Biochem. 189:73-81(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=90264400; PubMed=2160960;  
 RA Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;  
 RT "The complete sequence of botulinum neurotoxin type A and comparison  
 RT with other clostridial neurotoxins";  
 RL J. Biol. Chem. 265:9153-9158(1990).  
 RN [3]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding  
 RT components of the botulinum toxin complex in proteolytic Clostridium  
 RT botulinum types A, B, and F: evidence of chimeric sequences in the  
 RT gene encoding the nontoxic nonhemagglutinin component";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 RN [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RC STRAIN=Hall;  
 RX MEDLINE=89350959; PubMed=2669749;  
 RA Betley M.J., Somers E., Dasgupta B.R.;  
 RT "Characterization of botulinum type A neurotoxin gene: delineation of  
 RT the N-terminal encoding region";  
 RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).

[5]  
 RN SEQUENCE OF 1-18 FROM N.A.  
 RC STRAIN=Type A NIH;  
 RX MEDLINE=96096783; PubMed=8521962;  
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
 components of Clostridium botulinum type A progenitor toxins";  
 RL FEBS Lett. 376:41-44(1995).  
 RN [6]  
 RN SEQUENCE OF 1-16.  
 RX MEDLINE=84178501; PubMed=6370252;  
 RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;  
 RT "Partial amino acid sequence of the heavy and light chains of  
 RT botulinum neurotoxin type A";  
 RL Biochem. Biophys. Res. Commun. 119:900-904(1984).  
 RN [7]  
 RN SEQUENCE OF 1-46.  
 RX Dasgupta B.R., Foley J., Niece R.;  
 RT "Partial sequence of the light chain of botulinum neurotoxin type A";  
 RL Biochemistry 26:4162-4162(1987).  
 RN [8]  
 RN SEQUENCE OF 1-5 AND 44-456.  
 RX MEDLINE=91120847; PubMed=2126206;  
 RA Dasgupta B.R., Dekleva M.L.;  
 RT "Botulinum neurotoxin type A: sequence of amino acids at the  
 RT N-terminus and around the nicking site";  
 RL Biochimie 72:661-664(1990).  
 RN [9]  
 RN SEQUENCE OF 448-454 AND 872-895.  
 RX MEDLINE=89024662; PubMed=3178218;  
 RA Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;  
 RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two  
 RT halves and their partial sequences";  
 RL Arch. Biochem. Biophys. 266:142-151(1988).  
 RN [10]  
 RN SEQUENCE OF 448-482.  
 RX MEDLINE=85285016; PubMed=3896784;  
 RA Shone C.C., Hambleton P., Melling J.;  
 RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin  
 RT and purification of two tryptic fragments. Proteolytic action near  
 RT the COOH-terminus of the heavy subunit destroys toxin-binding  
 RT activity";  
 RL Eur. J. Biochem. 151:75-82(1985).  
 RN [11]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94063091; PubMed=8243676;  
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds";  
 RL FEBS Lett. 335:99-103(1993).  
 RN [12]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94124495; PubMed=8294407;  
 RA Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT "Proteolysis of SNAP-25 by types E and A botulin neurotoxins";  
 RL J. Biol. Chem. 269:1617-1620(1994).  
 RN [13]  
 RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.  
 RX MEDLINE=21556941; PubMed=11700044;  
 RA Rigoni M., Caccini P., Johnson E.A., Montecucco C., Rossetto O.;  
 RT "Site-directed mutagenesis identifies active-site residues of the  
 RT light chain of botulinum neurotoxin type a";  
 RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
 RX MEDLINE=98455071; PubMed=9783750;  
 RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;  
 RT "Crystal structure of botulinum neurotoxin type A and implications  
 RT for toxicity";  
 RL Nat. Struct. Biol. 5:898-902(1998).  
 CC !- FUNCTION: Inhibits acetylcholine release. The botulinum toxin



[illegible][illegible]

RESULT 6		STANDARD;	PRT;	1250 AA.
EXE_CLOBO	ID	BXE_CLOBO		
AC	000496;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)			
DE	(Bontoxiliysin E)			
OS	Clostridium botulinum.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
ON	NCBI_TaxID=1491;			
ON	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=Beluga;			
RX	MEDLINE=92181428; PubMed=1543481;			
RA	POulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;			
RT	Sequences of the botulin neurotoxin E derived from Clostridium			
RT	botulinum type E (strain Beluga) and Clostridium butyricum (strains			
RT	ATCC 43181 and ATCC 43755).;			
RL	Biophys. Res. Commun. 183:107-113(1992).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92174922; PubMed=1541280;			
RA	Whelan S.M., Elmcree M.J., Bodsworth N.J., Atkinson T., Minton N.P.;			
RT	The complete amino acid sequence of the Clostridium botulinum type-E			
RT	neurotoxin, derived by nucleotide-sequence analysis of the encoding			
RT	gene.;			
RL	Eur. J. Biochem. 204:657-667(1992).			
RL	[3]			
RP	SEQUENCE OF 1-251 FROM N.A.			



DR	InterPro:	IPR000130;	Zn_MtPeptidse.
DR	Fram:	Pf01742;	Peptidase_M27; 1.
DR	Prints:	PRO0760;	BONTXILYSIN.
DR	ProDom:	PD001963;	Bontoxilysin; 1.
DR	PROSITE:	PS00142;	ZINC_PROTEASE; 1.
KW	Neurotoxin;	Transmembrane;	Hydrolase; Metalloprotease; Zinc.
FT	INIT_MET	0	
FT	CHAIN	1	421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT	CHAIN	422	1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT	METAL	211	211 ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	212	212 BY SIMILARITY.
FT	METAL	215	215 ZINC (CATALYTIC) (BY SIMILARITY).
FT	DISULFID	411	425 INTERCHAIN (PROBABLE).
FT	CONFLICT	176	176 R -> G (IN REF. 2).
FT	CONFLICT	197	197 C -> S (IN REF. 2 AND 3).
FT	CONFLICT	339	339 R -> A (IN REF. 2).
FT	CONFLICT	772	772 I -> L (IN REF. 2).
FT	CONFLICT	962	962 EE -> QQ (IN REF. 2).
FT	CONFLICT	966	966 R -> A (IN REF. 2).
FT	CONFLICT	1194	1194 N -> NN (IN REF. 2).
SQ	SEQUENCE	1250 AA;	143712 MW; D9FCEDDDA041EB4 CRC64;

Query Match 24.4%; Score 583; DB 1; Length 1250;  
 Best Local similarity 31.9%; Pred. No. 1.7e-27;  
 Matches 148; Conservative 88; Mismatches 166; Indels 62; Gaps

QY	2	TIFNFTSTNSLKDIIENEFNNINDSKILSLONRNKTLVDISGYNAEVSERGDVQLN	61
DB	821	SIDFKLSSTYDOKILSYFNKFRTFFKRKKSSVLNMRYKNDKYVDTSYGDSININGDVTKY	880
QY	62	PIFPDFKLGGSEDRGKIVTVQENIVNVSNMYESFSISFWIRI----	NKVWSLPGYTI 117
DB	881	PTKNQF--GIYNDKLSEVNISQNDYIIVDNKYKNFSISFWVRIPNYDNKIYVNVNEVTI	938
QY	118	IDSVK--NNSGSWGIIISFLVTLKQNEDESGSINFSDISNNAPCY-NKWFVFVTVMNM	175
DB	939	INCMRDNNSGKVASLNHNHNIITFDNRGINOKLAFNPNGANGISDYINKKFIYVITNDR	998
QY	176	MGNKVIYINGKLIDTTIKVELTGINFSTITPEINKIPDTGLITSDSNINMWIRDFYIF	235
DB	999	LGDSKLYINGNLIDQKSILNLGNIHVSDNILEKIVNCSTRYI-----	GIRYPNF 1049
QY	236	AELDGKDINILFNSLOYIVVKVDWGNDLRNKYGVVYVNDIYLNRYM--YANSRQIVFN	293
DB	1050	DKELDETEQTLYSNBPNTNILKDWGNVLLYDKBYILLNLKPNNFIDRRKSDTSLSN	1109
QY	294	TRSN---NNDNFEGYKIIKRI-RGNTNTRYRGGDILYFDMTINNKAYNLFMKNETMYA	349
DB	1110	IRSTILLANRLXSGIKVKIQRVNNSSTDNLVRKNDQVYVNFVASKT-HLF----	PLVA 1163
QY	350	DNHSTEDIAIGREOTKDINDNIFIQOPMNNTYYASQIFKSFNGENISGCISGTY	409
DB	1164	DTATNKETIKISSGSRFNVQVVY-----NSVGNCMTMFKN--NGNN-----	IPLL 1210
QY	410	RFR-----LGSDWRHYRLVPTYKQGNYSALLESTTH---WGQV	446
DB	1211	GFKADTWASTWI-----YTHMRCHTNSNGCFWNFI	1241

RESULT 7  
 BXB\_CLOBO STANDARD; PRT; 1290 AA.  
 ID BXB\_CLOBO  
 AC F10844; P10843;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BONT/B)  
 DE (Bontoxilysin B).  
 GN BoTB.  
 OC Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 NCBI TaxId=1491;

RA RP SEQUENCE FROM N.A.  
 RX MEDLINE=92384550; PubMed=1514783;  
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,  
 RA Minton N.F.;  
 RT "Molecular cloning of the Clostridium botulinum structural gene  
 RT encoding the type B neurotoxin and determination of its entire  
 RT nucleotide sequence.";  
 RL Appl. Environ. Microbiol. 58:2345-2354(1992).  
 RN [2]  
 RN SEQUENCE OF 35-245 FROM N.A.  
 RC STRAIN=NCTC 7273;  
 RA Szabo B.A., Pemberton J.M., Desmarchellier P.M.;  
 RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE OF 633-993 FROM N.A.  
 RC STRAIN=NCTC 7273;  
 RX MEDLINE=94013372; PubMed=8408542;  
 RA Campbell K., East A.K., Collins M.D.;  
 RT "Gene probes for identification of the botulinum neurotoxin gene and  
 RT specific identification of neurotoxin types B, E, and F.";  
 RL J. Clin. Microbiol. 31:2255-2262(1993).  
 RN [4]  
 RN SEQUENCE OF 1-44 AND 441-466.  
 RC STRAIN=657;  
 RX MEDLINE=89000987; PubMed=3139097;  
 RA Dasgupta B.R., Datta A.;  
 RT "Botulinum neurotoxin type B (strain 657): partial sequence and  
 RT similarity with tetanus toxin.";  
 RL Biochimie 70:811-817(1988).  
 RN [5]  
 RN SEQUENCE OF 1-16 AND 441-458.  
 RC STRAIN=OKRA;  
 RX MEDLINE=85197963; PubMed=3888113;  
 RA Schmidt J.J., Sachyamoorthy V., Dasgupta B.R.;  
 RT "Partial amino acid sequences of botulinum neurotoxins types B and  
 RT E.";  
 RL Arch. Biochem. Biophys. 238:544-548(1985).  
 RN [6]  
 RN IDENTIFICATION AS ZINC-PROTEASE.  
 RX MEDLINE=93054694; PubMed=14296590;  
 RA Schlavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;  
 RT "Botulinum neurotoxins are zinc proteins.";  
 RL J. Biol. Chem. 267:23479-23483(1992).  
 RN [7]  
 RN IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=93063293; PubMed=1331807;  
 RA Dasgupta B., Benfenati F., Poulin B., Rossetto O., de Laureto P.P.,  
 RA Schlavo G., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 RT by proteolytic cleavage of synaptobrevin.";  
 RL Nature 359:832-835(1992).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-1-PHE-77 BOND OF  
 CC SYNAPTOSOMALIN-2.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIANE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81186; AAA23211.1; -;  
 DR EMBL; Z11934; CAA77991.1; -;  
 DR EMBL; X70817; CAA50148.1; -;  
 DR PIR; S07128; S07128;  
 DR PIR; S07155; S07155;  
 DR PIR; S08562; S08562;  
 DR PIR; S08573; S08573;  
 DR PIR; S08574; S08574;  
 DR PIR; A48940; A48940;  
 DR HSSP; P10845; 3BFA;  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_MTPeptidse.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT\_MET 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 436 445 INTERCHAIN (PROBABLE).  
 FT DISULFID 29 29 T -> M (IN REF. 4).  
 FT CONFLICT 217 217 R -> G (IN REF. 2).  
 FT CONFLICT 224 224 A -> S (IN REF. 2).  
 FT CONFLICT 463 463 S -> R (IN REF. 4).  
 SQ SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;  
  
 Query Match 24.3%; Score 581; DB 1; Length 1290;  
 Best Local Similarity 32.9%; Pred. No. 2.3e-27;  
 Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;  
  
 QY 3 IPFNFSYTNNSLLKDIINEYFNINDSKILSNRKNLTVDTSGYNAEVSEGDVQLNP 62  
 Db 835 MPFOLSIYNTDILIEFMFNKYNSEILNLIILRYKONNLDLSGYCAKVEVDGVELND 894  
  
 QY 63 IFFPDFKLGSSEDRGKVIYTONENIVNYSFISFWIRINKW---VSNL--PGYT 116  
 Db 895 --KNQFKLTSSA--NSKIRVTQNNIIFNSVFLDFSVFWIRIPKYNKNCIQNYIINEFT 950  
  
 QY 117 IIDSVKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNNAPGY-NKWFVFTVNNM 175  
 Db 951 IINCNNKNSGWSKISGRNRIITWTLIDINGKTSVFVEYNIREDISEYINRWFVTITNN- 1009  
  
 QY 176 MGNMKIYINGKLIDITIKVKELTGINFSGKTIFFELINKIPDGLITSDSDNIN-MWIRDFYI 234  
 Db 1010 LNNAKIYINGKLENSWDIKDIREVANGELIFKLD-----GIDRTQCFIWMKYFSI 1060  
  
 QY 235 FAKELDGKDINILFNSLQYTNVVKDYNGNDRYNKEYVMYNDILNRYMYANS-----RQI 290  
 Db 1061 FNTLESQSNIEERYKIQSYSEYKDFWGNPLNKEYYMFNAGNKNSYIKLKDSDPVGEI 1120  
  
 QY 291 VENTERNNN-----DENEGYKIIIRKRGNT---NDTRVRGBDILYFD-MTINKK--- 336  
 Db 1121 ITRSKYNSKYNINRYDLYIGKEFIIRR-XSNSQSINDIVRKEDYIYLDFFNLNDEWRV 1179  
  
 QY 337 -AYNLFMKNETMYADNHSTEDIYAIGREOTKDINDNIIIFIQPMNNTYVYASQ-IFKSN 394  
 Db 1180 YTYKYFKKEE-----EKLFLAPISDSDEFYN--TQIKEYDEQPYSQLLEKK- 1226  
  
 QY 395 FNGENISGICIGYRFLGQ-----DWTNRHNYLVTYKQGNYSALLSESTST 441  
 Db 1227 -DEESTDEIGLIGHRFVESGIVPEEYKDFYKISKWY-----LKEVKKRPYNLKL---GC 1277

Qy 442 HWGFVPVSE 450  
 Db 1278 NWQFIPKDE 1286

RESULT 8  
 BXA2\_CLOBO STANDARD; PRT; 1295 AA.  
 AC Q45894; P77780;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)  
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].  
 GN BOTA OR BNA OR ATX  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Kyoto-F;  
 RX MEDLINE=94143603; PubMed=8310180;  
 RA Williams A., East A.K., Lawson P.A., Collins M.D.;  
 RT "Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins.";  
 RL Res. Microbiol. 144:547-556(1993).  
 RN [2]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN=Kyoto-F;  
 RX MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component.";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 CC -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-1-Arg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.  
 CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H) (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL; X73423; CAA51824.1; -;  
 DR EMBL; X87974; CAA61234.1; -;  
 DR HSSP; P10845; 3BTA.  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.

DR InterPro; IPR000130; zn\_MTPeptidse.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; Hydrolase; Metalloprotease; Zinc.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT\_MET 0 0  
 FT CHAIN 1 447 BY SIMILARITY.  
 FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.  
 FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 223 223 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 226 226 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 429 433 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 1234 1279 BY SIMILARITY.  
 FT TRANSMEM 626 646 POTENTIAL.  
 FT TRANSMEM 655 675 POTENTIAL.  
 SQ SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

Query Match 23.2%; Score 554.5; DB 1; Length 1295;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-26;  
 Matches 147; Conservative 81; Mismatches 183; Indels 67; Gaps 12;

Qy 3 IPFNIPSYTNNLLKDIINEYFNINDSKILSLQNKNTLVDTSGYNAEVSEEGDVQLNP 62  
 Db 848 IPQLSKYVDNKKLLSTFTYIKNIWNTSILSVYKDDLDLSRYGAKINIGRVYDS 907  
 Qy 63 IFPPDFKLSSGDEGRKVIQVTONENIVNMYESFSISFWIRNKWVSNL---POYTIID 119  
 Db 908 IDKNOIKLINLESSTIEVIL--KNAIVNMYENFSTFWIKPKYFSKINLNNEYTIIN 965  
 Qy 120 SVKNNSGWSGIISNPLVFTLKNQEDSEOSINFSYDISNNAPGY-NKAPFFVTVTNNMGN 178  
 Db 966 CIENNSGKVSINLYGELIITWLNQNKONIQRVVFKISQWYNSIDYINRWITVTITNRLTK 1025  
 Qy 179 MKIYINGKLDTIKVKELTGINFSTKITTEINKIPDITGLTSDSNINNMWIDFYFAKE 238  
 Db 1026 SKIYINGRLIDQKIPISNLGNIHASNKIMEKDCRD-----PRRYIMIKYFNLFOKE 1077  
 Qy 239 LDGKIDILNLSQYINNVKDYNGNDLRYNKKEYMYNIDYLNRYMANS----- 287  
 Db 1078 LNEKEIKDLYDSOSNGILKDFWGNLYQDKPYIMUNLFPKRYVDVNNIGRGWYLGK 1137  
 Qy 288 --RGIVFTRNNNDNFEGYKIIKIRIGTNDTRVGGDILYFDMTINNKAYNLFMKNE 345  
 Db 1138 PRGSVVTTIYNLSTLYEGTKFIKKYASGNEINVRNDRVYINVVVKNEYRL----- 1192  
 Qy 346 TMADNHSTEDIYVGLRGQTKDINDINIFQI-----QPMNNTYYVASQIFKSNFG 397  
 Db 1193 ---ATNASQAGVEKILSALPIPDVGN--LSQVVVMKSKDDQGIKRN-----CKMNLQD 1240  
 Qy 398 ENISGICISGYTRF-----RLGGDWYRHNLYVPTVQKNYASLLESTSTHWGFVPVSE 450  
 Db 1241 NNGNDIGFIFGLHYDNLAKLVASNNYN-----RQVGKAS--RTFGCSWEFIPVDD 1288

RESULT 9  
 BXA2\_CLOBO STANDARD; PRT; 1274 AA.  
 AC P30936;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BONT/F)  
 DE (Bontoxilysin F).  
 GN BOTF.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 23387;  
 RX MEDLINE=93012902; PubMed=1398040;  
 RA East A.K., Richardson P.T., Allaway D., Collins M.D.,

Roberts T.A., Thompson D.E.;  
"Sequence of the gene encoding type F neurotoxin of Clostridium botulinum";  
FEMS Microbiol. Lett. 75:225-230(1992).  
[2]  
SEQUENCE OF 1-64 FROM N.A.  
STRAIN-Hobbs FT10;  
MEDLINE=94297488; PubMed=7764998;  
RA East A.K., Collins M.D.;  
"Conserved structure of genes encoding components of botulinum neurotoxin complex M and the sequence of the gene coding for the nontoxic component in nonproteolytic Clostridium botulinum type F.";  
Curr. Microbiol. 29:69-77(1994).  
[3]  
SEQUENCE OF 634-1002 FROM N.A.  
MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
"Gene probes for identification of the botulin neurotoxin gene and specific identification of neurotoxin types B, E, and F";  
J. Clin. Microbiol. 31:2255-2262(1993).  
[4]  
IDENTIFICATION OF SUBSTRATES.  
RA MEDLINE=94230352; PubMed=8175689;  
RA Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,  
Romes B., Fyke E.M., Suedhof T.C., Jahn R., Niemann H.;  
"Cleavage of members of the synaptobrevin/VAMP family by types D and F botulin neurotoxins and tetanus toxin";  
J. Biol. Chem. 269:12764-12772(1994).  
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-LYS-59 BOND OF SYNAPTOSREVINS-1 AND -2.  
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No substrate action on small molecule substrates.  
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC -1- SUBCELLULAR LOCATION: Secreted  
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M92906; AAA23263.1; -  
CC EMBL: S73676; AAC60475.1; -  
CC EMBL: X70820; CAA50151.1; -  
CC EMBL: X70816; CAA50147.1; -  
CC HSSP: P10845; 3BFA.  
CC  
CC MEROPS: M27.002; -  
CC InterPro: IPR000395; Bontoxilysin.  
CC InterPro: IPR000130; Zn\_Mtpeptidse.  
CC Pfam: PF01742; Peptidase\_M27; 1.  
CC PRINTS: PR00760; BONTOXILYSIN.  
CC ProDom: PD001963; Bontoxilysin; 1.  
CC PROSITE: PS00142; ZINC\_PROTEASE; 1.  
CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
FT CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.  
FT CHAIN 237 1274 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.  
FT METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 228 228 BY SIMILARITY.  
FT METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 231 231

RT fragment C in *Escherichia coli*.  
 RL J. Bacteriol. 165:21-27(1988).  
 RN [4]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE-90201034; PubMed-2108021;  
 RA Krieglstein K., Hubschen A., Weller U., Habermann E.;  
 RT "Arrangement of disulfide bridges and positions of sulphydryl groups  
 in tetanus toxin."  
 RL Eur. J. Biochem. 188:39-45(1990).  
 RN [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE-92037649; PubMed-1935979;  
 RA Krieglstein K.G., Hubschen A.E., Weller U., Habermann E.;  
 RT "Limited proteolysis of tetanus toxin. Relation to activity and  
 identification of cleavage sites."  
 RL Eur. J. Biochem. 202:41-51(1991).  
 RN [6]  
 RP IDENTIFICATION AS ZINC-PROTEASE.  
 RX MEDLINE-93010948; PubMed-1396558;  
 RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,  
 RA Montecucco C.;  
 RT "Tetanus toxin is a zinc protein and its inhibition of  
 neurotransmitter release and protease activity depend on zinc."  
 RL EMBO J. 11:3577-3583(1992).  
 RN [7]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE-93063293; PubMed-1331807;  
 RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RA Dasgupta B.R., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 by proteolytic cleavage of synaptobrevin."  
 RL Nature 359:832-835(1992).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.  
 RX MEDLINE-97475217; PubMed-9334741;  
 RA Unland T.C., Wiegert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,  
 RA Sax M.;  
 RT "Structure of the receptor binding fragment HC of tetanus  
 neurotoxin."  
 RL Nat. Struct. Biol. 4:788-792(1997).  
 CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-1-PHE-77  
 BOND OF SYNAPTOSOMAL-2.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-1-PHE-77 BOND IN  
 SYNAPTOSOMAL-2.  
 CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO  
 YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE  
 AND ARE NON-TOXIC AFTER SEPARATION.  
 CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO  
 GANGLIOSIDE RECEPTORS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X04436; CAA28033.1; -  
 CC EMBL: M12739; AAA23282.1; -  
 CC EMBL: X06214; CAA29564.1; -  
 CC PIR: A25689; BTCLTN  
 CC PDB: 1A29; 29-APR-98.  
 CC PDB: 1A80; 14-OCT-98.  
 CC MEROPS: M27.001; -  
 CC InterPro: IPR000395; Bontoxilysin.  
 CC InterPro: IPR000130; Zn\_Mtpeptidase.

DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINIS; PRO0760; Bontoxilysin.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;  
 KW 3D-structure.  
 FT INIT\_MET 0  
 FT CHAIN 1 456  
 FT CHAIN 457 1314  
 FT METAL 232 233  
 FT ACT\_SITE 233 233  
 FT METAL 236 236  
 FT TRANSMEM 226 246  
 FT TRANSMEM 669 689  
 FT DISULFID 438 466  
 FT DISULFID 1076 1092  
 SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EP81D CRC64;  
 Query Match 17.8%; Score 425; DB 1; Length 1314;  
 Best Local Similarity 26.4%; Pred. NO. 4.5e-18;  
 Matches 134; Conservative 90; Mismatches 169; Indels 114; Gaps 21;  
 QY 3 IPP----NIFSYNNLSLLKDIINEYFNNINDSKILQNRKNTLVDTSGYNAEYSEGDV 58  
 DB 858 IPFSYKSLDCWVDNEEDIVI-----LKRSTILNLDINNDISDIFGNSVITYPDA 911  
 QY 59 QLNPIPFDFKLGSSCE-----DRGKVIYVTONENIVYNSMYESFISFWIRNK-WVS 110  
 DB 912 QLVP-----GINKATHLVNSESSEVIVHKAMDIEYDNMFNFTVFWLRVPKVSAS 963  
 QY 111 NLP-----GYTIIDSYKNN-----SWSIGIISNLFVTLKQNEDESGSINFSYDISNNA 160  
 DB 964 HLEQYGTNEYSISSMKHSLSGSGWSVSLKGNLTLTKDSAGEVQITFR-DLPDKF 1022  
 QY 161 PGV--NKWFEVTVNNMGNMKIYVINGKLIDIKVKELTGINFSTKITFEINKIPDTGLI 218  
 DB 1023 NAYLANKWFIITNDRLSSANLYINGVLMSAEITGLGAIREDDNNITLKDR----- 1075  
 QY 219 TSQSDNINWIRDFYFAKELGDKDINILFNSLQTVNVVQYWGNDLRYKEYVMNI-- 276  
 DB 1076 -CNQYYSIDKFRIFCALNPKIEKLYTSITTFIRDFWGNPLRYDTEYLLIPVAS 1134  
 QY 277 -----DYL--NRYMANGROIVENRRNNDFNCEYKIIKIRGNTN-DTRV 321  
 DB 1135 SSKDVLKNTIDYMLTNAPSVYNGKLIYRYLYN-----GLAFIKRYTPNEIDSFV 1189  
 QY 322 RGGDILYFDMTNNKAY-----NLFMKNMTYADNHSTEDIYAGLREOTKDINDNI 373  
 DB 1190 KSGDFIKLYSVYNNNEHIVGPKDGNFNNLDRILRVGYNAPGIPLYKKMEAVK----- 1243  
 QY 374 IFQIQPMNNTYYVASQIFKSNFNGENISGICSGIYGRRLGSD-----WYRHYL 423  
 DB 1244 -----LRDLKTYSVOL--KLYDDKNAS-LGLVGHNGICGNDPNDRIILIASNWY-FNHL 1293  
 QY 424 VPIVKGNYASLLESTSTHWGFVPVSE 450  
 DB 1294 KDXI-----LGCDWYFVPTDE 1309  
 RESULT 11  
 EXEN\_CLOBO STANDARD; PRT; 1162 AA.  
 AC P46082;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Botulinum neurotoxin type E, nontoxic component.  
 GN ENT-120.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=Washington;
RX MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
RL J. Gen. Microbiol. 139:79-86(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12697; BAA02194.1;
DR InterPro: IPR000395; Bontotoxylisin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR ProDom; PD001963; Bontotoxylisin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDAAE0F39D CRC64;
Query Match 9.6%; Score 230.5; DB 1; Length 1162;
Best Local Similarity 22.8%; Pred. No. 1.5e-06;
Matches 96; Conservative 74; Mismatches 158; Indels 93; Gaps 17;
QY 6 NIFSNTNSLLKDIINEYFNNDKILSKLQNRKNTLV-DTSGYNAEVSSEGQVQLNP 64
DQ 783 NLFNLYTELLIKEQTSFY-----ELSLYAFQEQDNNVIGDTSKNTLVE-----Y 827
QY 65 PDFKLGSSGDEGRKVIYQNTENIVY-----NSMYSFSTSFIRINKWVSNLPGYTI- 117
DQ 828 PKDIGLVYGINNATHLTGANOIKFTNDYFENGLTNFSLYFWR-----NLQNTIK 881
QY 118 ---IDSVKNNSGWISGIIISNPLVFTLKQNEDEQSINFSYDINSNAPGYNKFFVTVT-N 173
DQ 882 SKLIGSKEDNCGWEIFYFENGLVFNIIIDSGNEKNIVLS-NISNKS-----WHYIVISIN 935
QY 174 NMGMNKIYINGKLIDTIVKVELGINSKITTEINKIPDTGLTSDSNINWIRDFY 233
DQ 936 RLKQDQLLIFIDNVLVANEDEIKELINYS-----DIISLSDNNNVYIEGLS 982
QY 234 IFAKELDGKIDNILENSLYQYNNVYKDWGNDLYNKYVWVNYD-----LNRYMYANSRQ 289
DQ 983 VLNKTINSNEILTDYFSLNNSYIRNFDEEILQYRNYELFNVYFPEIAINKIEQNNYI 1042
QY 290 IVFNTRNNDFNEGYKIIIRKIRGNTNDR-----VRGGDILYFDMTINNK 337
DQ 1043 LSTNNENLNFKPLKFKLL-----NTNPKQYQVKWDEVIFSLDTEKYLDISTIN-- 1094
QY 338 YNLFMKNMETHYADNHSTEDIIAIGLRQETKDINDNIIFQIQPM--NNTYYVASOIFKSNF 395
DQ 1095 -----NRQLVDNKNNAQIFIN-----NDIFISNCLITLYNNVNVYLS-IKNQDY 1139
QY 396 N 396
DQ 1140 N 1140

```

## RESULT 12

```

BXCN_CLOBU
ID BXCN_CLOBU STANDARD; PRT; 1162 AA.
AC Q06366;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```

```

CC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=93360835; PubMed=8355622;
RA Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RT "Similarity in nucleotide sequence of the gene encoding nontoxic
RL component of botulinum toxin produced by toxigenic Clostridium
RT butyricum strain BL6340 and Clostridium botulinum type E strain
RT Mashike.";
RL Microbiol. Immunol. 37:395-398(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12739; BAA02231.1;
DR InterPro: IPR000395; Bontotoxylisin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR ProDom; PD001963; Bontotoxylisin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136829 MW; C86E9BE932DA78E4 CRC64;
Query Match 9.6%; Score 230.5; DB 1; Length 1162;
Best Local Similarity 22.9%; Pred. No. 1.5e-06;
Matches 97; Conservative 76; Mismatches 152; Indels 99; Gaps 19;
QY 6 NIFSNTNSLLKDIINEYFNNDKILSKLQNRKNTLV-DTSGYNAEVSSEGQVQLNP 64
DQ 783 NLFNLYTELLIKEQTSFY-----ELSLYAFQEQDNNVIGDTSKNTLVE-----Y 827
QY 65 PDFKLGSSGDEGRKVIYQNTENIVY-----NSMYSFSTSFIRINKWVSNLPGYTI- 117
DQ 828 PKDIGLVYGINNATHLTGANOIKFTNDYFENGLTNFSLYFWR-----NLQNTIK 881
QY 118 ---IDSVKNNSGWISGIIISNPLVFTLKQNEDEQSINFSYDINSNAPGYNKFFVTVT-N 173
DQ 882 SKLIGSKEDNCGWEIFYFENGLVFNIIIDSGNEKNIVLS-NISNKS-----WHYIVISIN 935
QY 174 NMGMNKIYINGKLIDTIVKVELGINSKITTEINKIPDTGLTSDSNINWIRDFY 233
DQ 936 RLKQDQLLIFIDNVLVANEDEIKELINYS-----DIISLSDNNNVYIEGLS 982
QY 234 IFAKELDGKIDNILENSLYQYNNVYKDWGNDLYNKYVWVNYD-----LNRYMYANSRQ 289
DQ 983 VLNKTINSNEILTDYFSLNNSYIRNFDEEILQYRNYELFNVYFPEIAINKIEQNNYI 1040
QY 290 IVFNTRNNDFNEGYKIIIRKIRGNTNDR-----VRGGDILYFDMTIN 334
DQ 1041 -IYLSNENLNFKPLKFKLL-----NTNPKQYQVKWDEVIFSLDTEKYLDISID 1093
QY 335 NKAYNLFMKNMETHYADNHSTEDIIAIGLRQETKDINDNIIFQIQPM--NNTYYVASOIFK 392
DQ 1094 N-----NRQLVDNKNNAKTFIN-----NDIFISNCLITLYNNVNVYLS-IKN 1136
QY 393 SNFN 396
DQ 1137 QDYN 1140

```

## RESULT 13

```

BXCN_CLOBU
ID BXCN_CLOBU STANDARD; PRT; 1196 AA.
AC P46081;

```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type C1, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=92231894; PubMed=1567404;
RA Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Ogura K.;
RT "The complete nucleotide sequence of the gene coding for the
RT nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT progenitor toxin."
RL Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62389; CAA44262.1;
DR Interpro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
KW
SQ SEQUENCE 1196 AA; 138740 MW; 4BD5956274D7F9C3 CRC64;
Query Match 8.3%; Score 199; DB 1; Length 1196;
Best Local Similarity 18.8%; Pred. No. 0.00011;
Matches 86; Conservative 89; Mismatches 151; Indels 132; Gaps 18;
QY 6 NIFSNTNSLLDIINEYFNNDKSLSLONR---KNTLVDTSGYNAEYSEEGDVOLNP 62
DB 818 NFNSOVEQVMEILSPY-----QLLFASKGPNISNIEDISGKNTLQYTESIEL-- 868
QY 63 IPFDFKLGSSGEDRGKIVIVTQNEIVY-----NSMYESFSISFWIRINKWNSLPGYT 116
DB 869 -----VYGNGE--SLYLKSPNETIKFSNFFFTNGLTNNETICFWLRTG--KNDDKTR 918
QY 117 IIDSVKNNSGWSIGIISNLFVTLKQNESQSINFSDYDISNAPGYNKQFFVTVT--NNM 175
DB 919 LIGNKVNCGWEIYFEDNGLVFEIIDSNGQESVYLSINDN-----WYISISVDRL 972
QY 176 MGNMKIYINGKLIDTIKYKELTGINFSTIITFEINKIPDGTGLTSDSDNINMWIRDFYIF 235
DB 973 KDQLLIFINDKNVANSIDQILSI-YSTNIISLVNK-----NNSIYVEELSVL 1019
QY 236 AKELDGKDNILFNSLQTVNVKDWGNDLRNKEYYVMNIDYLRNRYANSRQIVENTR 295
DB 1020 DNPIITSEEVIRNYFSLDYSYIRDSKSLLEYKNYQLYINRVFFETSLY----- 1068
QY 296 RNNNDNFEGYKIIIRKIRGNINDTRVRGGDILYFDMTLNKAYLNFEMNRYADNHS 355
DB 1069 -----EYND-----NNSY-LSLKN-----TD 1084
QY 356 DYIAYGLRQTKDINDNIF-----OTOPNNNTYYVASQIFKSNFN 396
DB 1085 GINISSVFKFLINIDESKYVQKWDECICVLDGTEKYLDISPENNR-----QLVSSKDN 1140
QY 397 GENISGICSIGHYRFLGSDWYRHN--YLVPTVKOGNY 432
DB 1141 AKKI-----IVNTDLRPPDCITFSYNDKVFSLRGGDY 1174
RESULT 14
RMAR_YEAST STANDARD; PRT; 396 AA.
ID P02381;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Mitochondrial ribosomal protein VAR1.
GN VAR1.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83050946; PubMed=6754087;
RA Hudspeth M.E.S., Ainley W.M., Shumard D.S., Butow R.A., Grossman L.I.;
RT "Location and structure of the var1 gene on yeast mitochondrial DNA:
RT nucleotide sequence of the 40.0 allele."
RT Cell 30:617-626(1982).
RL -!- FUNCTION: ESSENTIAL FOR MITOCHONDRIAL PROTEIN SYNTHESIS AND
RL REQUIRED FOR THE MATURATION OF SMALL RIBOSOMAL SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: BELONGS TO THE VAR1 FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V00705; CAA24077.1;
DR FIR; A02751; R3BIM1.
DR SGD; S0007275; VAR1.
KW Ribosomal protein; Mitochondrion.
KW
SQ SEQUENCE 396 AA; 46881 MW; 62A68A24740BCBE0 CRC64;
Query Match 6.9%; Score 164.5; DB 1; Length 396;
Best Local Similarity 22.8%; Pred. No. 0.0036;
Matches 97; Conservative 78; Mismatches 154; Indels 97; Gaps 23;
QY 12 NNSLKDIIINEYFNNDKSLSLONRKNLTVDTSYNAEYSEE--GDVQLNP--IPFPD 67
DB 19 NNNI---IINTLDSLMMKKLL-----LKNMLDMNKKMNNKMLNNNNNPPAGANPYV 71
QY 68 FKLGGSGEDRGKV-----IVTQNEI-VYKSMKESFSISFWIRINKWNSLPG 114
DB 72 HRIGPAGINNNKQLHNNNNNNNTQIYNNKMEIMTMNDKL-----INKLL----- 119
QY 115 YTIIDSVKNNSGWSIGIISNLFVTLKQNESQSINFY---DISNAPGYNKQFFVTY 171
DB 120 YKMWTLKLNMMNINKIIMSK-----TIHQSLNKLNIKIFYNNNDINN--NNNNNYMMN 173
QY 172 TNMMGNKIYINGKLIDTIKYKELTGINFSTIITFEINKIPDGTGLTSDSDNINMWIR 231
DB 174 MKKLNIIMNNNNNLCN-----ILSYNNKVTIEPIKL---SYILNSD-----IFS 219
QY 232 FYIFAKELDGKDINLFN-SLOYTNVVKDWGNDLRNKEYYVMNIDYLRNRYANSRQI 290
DB 220 KYLSLNDMDKYNGGLTNYQRLNNIPKLNHDNISMN---YINNINNNNNKNNINL 276
QY 291 VENTR--RNNNDNFEGYKIIIRKIRGNINDTRVRGGDILYFDMTINNKAYNLFM----- 342
DB 277 LNNNNNNNNNNNNN-----NNNNNYIGNNNIYNNNTIDNPMDILMYKYLVG 324
QY 343 ---KNETVADNH---STEDIYVAILGREQTK---DINDNIIFQIOPMNNNTYYVASQIFK 392
DB 325 WSIKFGKRLSNNNGTSTNNLN-GTFNNKAYLWSNNNNKLNIFSPSNLYANSNI-- 361
QY 393 SNFNGE 398
```

382 - NKNGK 386

RESULT 15

```

-----
C7AB_BACUK          STANDARD;          PET;      1138 AA.
ID C7AB_BACUK
AC Q45708;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidial crystal protein cry7AB (Insecticidal delta-endotoxin
DE cryVIIA(b)) (Crystalline entomocidal protoxin) (130 kDa crystal
DE protein).
DE CRY7AB OR CRYVIIA(B).
GN Bacillus thuringiensis (subsp. kumamotoensis).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=132267;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=HD867;
RC
RT Payne J.M., Fu J.M.;
RA "Coleopteran-active Bacillus thuringiensis isolates and genes encoding
RA coleopteran-active toxins.";
RL Patent number US286486, 15-FEB-1994.
RL -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF COLEOPTERA.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@sb-sib.ch).
-----
EMBL; U04368; AAA21121.1;
DR HSSP; P07130; 1DIC.
DR InterPro; IPR001178; Endotoxin.
DR Pfam; PF00555; endotoxin; 1.
DR Toxin; Sporulation.
SQ SEQUENCE 1138 AA; 129658 MW; E12DC80C0A56D1DA CRG54;
-----

```

QY	259	DYWGNDLRNKEYTWN-IDYLRYMYANSRQIVEN---TRRNNDFNEG----KIIIR 311
Db	681	DCISGLLPNEKREQLNDVYAKLSV---SRNLLDPFDPSINSENGWTSNGVIGN 738
QY	312	----IRGN-----TNDTRVGGDILYFDMTNNKAYNLFMKNETMYADN---HSTEDIY 358
Db	739	GDVFVKGNLYIFSSTNTQ-----YPTLYQKIDESKKKEYTRYKLGFISSQDL 790
QY	359	ATGLR-----EQTKDINDNIIFQIPMNNVTYYASGTFKSNFNGENIGSICS 405
Db	791	AVVIRYDAKHTLDVSNLLPDLIP-ENTCGPNRCAAQVLDENPSSQCS 840

Search completed: November 7, 2002, 14:44:02  
Job time : 34 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:40:06 ; Search time 81 Seconds

(without alignments)  
1144.706 Million cell updates/sec

Title: US-09-910-186a-10  
Perfect score: 2391  
Sequence: 1 MTIFNFSYNNSLKDI.....NYASLLESTHNGFVPSVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2386	99.8	1291	2 Q93HT3	Q93ht3 clostridium
2	1824	76.3	1285	2 Q9LBR1	Q9lbr1 clostridium
3	1823	76.2	1285	2 Q45B67	Q45b67 clostridium
4	882.5	36.9	1280	2 Q9LBS7	Q9lbs7 clostridium
5	882.5	36.9	1280	2 Q45849	Q45849 clostridium
6	867.5	36.3	1275	12 Q9QMG7	Q9qmg7 clostridium
7	610	25.5	1291	2 Q933K0	Q933k0 clostridium
8	604	25.3	1291	2 Q9ZAJ8	Q9zaj8 clostridium
9	602.5	25.2	1255	2 Q9FAR6	Q9far6 clostridium
10	602	25.2	1291	2 Q93G71	Q93g71 clostridium
11	594	24.8	1251	2 Q9K395	Q9k395 clostridium
12	593	24.8	1291	2 Q98077	Q98077 clostridium
13	575	24.0	441	2 Q9X708	Q9x708 clostridium
14	569	23.8	1268	2 Q45851	Q45851 clostridium
15	564	23.6	1278	2 Q97236	Q97236 clostridium
16	527.5	22.1	1280	2 Q9ZAJ5	Q9zaj5 clostridium

17	428	17.9	451	2	Q9LAI3	Q9lai3 clostridium
18	428	17.9	1310	2	Q93N27	Q93n27 clostridium
19	314	13.1	361	2	Q45848	Q45848 clostridium
20	312	13.0	361	2	Q45846	Q45846 clostridium
21	281.5	11.8	367	2	Q45861	Q45861 clostridium
22	281.5	11.8	367	2	Q45862	Q45862 clostridium
23	258	10.8	1197	2	Q45888	Q45888 clostridium
24	239.5	10.0	1198	2	Q06018	Q06018 clostridium
25	236.5	9.9	1197	2	Q33871	Q33871 clostridium
26	236.5	9.9	1197	2	Q9ZAJ9	Q9zaj9 clostridium
27	234	9.8	1197	2	Q9ZAJ9	Q9zaj9 clostridium
28	231.5	9.7	1198	2	Q69277	Q69277 clostridium
29	230.5	9.6	1198	2	Q45893	Q45893 clostridium
30	216.5	9.1	1193	2	Q45880	Q45880 clostridium
31	215.5	9.0	1161	2	Q45891	Q45891 clostridium
32	215.5	9.0	1161	2	Q69276	Q69276 clostridium
33	213	8.9	1193	2	Q45914	Q45914 clostridium
34	213	8.9	1193	2	Q9L107	Q9l107 clostridium
35	208	8.7	1162	2	Q9ZAJ6	Q9zaj6 clostridium
36	200	8.4	1196	2	Q9LBS8	Q9lbs8 clostridium
37	199	8.3	1165	2	Q45844	Q45844 clostridium
38	199	8.3	1165	2	Q45887	Q45887 clostridium
39	199	8.3	1196	2	Q45916	Q45916 clostridium
40	199	8.3	1196	2	Q93HT4	Q93ht4 clostridium
41	199	8.3	1196	9	Q38197	Q38197 clostridium
42	197	8.2	1163	2	Q45850	Q45850 clostridium
43	197	8.2	1196	2	Q9LBR2	Q9lbr2 clostridium
44	197	8.2	1196	2	Q93550	Q93550 clostridium
45	197	8.2	1196	9	Q9ZX77	Q9zx77 clostridium

#### ALIGNMENTS

RESULT 1  
Q93HT3 ID Q93HT3 PRELIMINARY; PRT; 1291 AA.  
AC Q93HT3;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Neurotoxin.  
GN Nr.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1491;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-C-YOICHI;  
RX MEDLINE=21534265; PubMed=11676492;  
RA Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,  
FUJINAGA Y., OGUMA K., OHYAMA T.;  
RT "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in  
RT Hemagglutination";  
RL Biochem. Biophys. Res. Commun. 288:650-657(2001).  
DR EMBL; AB061780; BABY1749.1; .  
DR InterPro; IPR000395; Bontoxilysin.  
DR Pfam; IPR000130; Zn\_Mrptdse.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR ProDom; PD001963; Bontoxilysin; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1  
SQ SEQUENCE 1291 AA; 148869 MW; 4A21DB35B8743CF8 CRC64;  
Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred.No. 5.7e-117;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TIFNFSYNNSLKDIINEYFNNDKILSKLNKNTLVDTSGYNAEVSSEGDVQLN 61  
Db 843 TIFNFSYNNSLKDIINEYFNNDKILSKLNKNTLVDTSGYNAEVSSEGDVQLN 902  
QY 62 PIFPFDFKLGSSGDEGRGVIVTQNEINIVNMYSEFSISFWIRNKWVNLPGYTIIDS 121

```

Db 903 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 962
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYKWFVTVTNMGMNMI 181
Db 963 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYKWFVTVTNMGMNMI 1022
Qy 182 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 241
Db 1023 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 1082
Qy 242 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYTANRSQIVFTRNNNDP 301
Db 1083 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYTANRSQIVFTRNNNDP 1142
Qy 302 NEGKIIIRKIRGTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHTEDIYAIG 361
Db 1143 NEGKIIIRKIRGTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHTEDIYAIG 1202
Qy 362 LREQTKDINDNIIFQIQPMNTYYASQIFKSNFNGENISGICSTGYRFRJGGDWYRHN 421
Db 1203 LREQTKDINDNIIFQIQPMNTYYASQIFKSNFNGENISGICSTGYRFRJGGDWYRHN 1262
Qy 422 YLVPVTKQGNYSALLESTSTHWGFVPVSE 450
Db 1263 YLVPVTKQGNYSALLESTSTHWGFVPVSE 1291

RESULT 2
Q9LBR1
ID Q9LBR1 PRELIMINARY; PRT: 1285 AA.
AC Q9LBR1
DT 01-OCT-2000 (Tremblrel. 15, Created)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D-4947;
RA Sagene Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT "Characterization of the progenitor toxin components produced by
RT Clostridium botulinum Type D Strain 4947";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB037920; BAA90861.1; -
DR HSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1285 AA; 147352 MW; B63AFA487D570680 CRC64;

Query Match 76.3%; Score 1824; DB 2; Length 1285;
Best Local Similarity 77.9%; Pred. No. 13e-87;
Matches 352; Conservative 36; Mismatches 56; Indels 8; Gaps 4;

Qy 2 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGQVQLN 61
Db 839 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGQVQLN 898
Qy 62 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 121
Db 899 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 958
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYKWFVTVTNMGMNMI 181

```

```

Db 959 KNSGWSIGIISNLFVTLKQNESEQDINFSYDISKNAAGYNKWFVTVTNMGMNMI 1018
Qy 182 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 1078
Qy 242 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYTANRSQIVFTRNNNDP 301
Db 1079 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYTANRSQIVFTRNNNDP 1138
Qy 302 NEGKIIIRKIRGTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHTEDIYAIG 360
Db 1139 NEGKIIIRKIRGTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHTEDIYAIG 1193
Qy 361 GLRQTKD-INDNIIFQIQPMNTYYASQIFKSNFNGENISGICSTGYRFRJGGDWY 418
Db 1194 GALDQPMDEIRKYGSGFIQPCNFDYASQLFLSSNATNRLGILSISYKFLGDDYWF 1253
Qy 419 RHNLYVTVKQGNYSALLESTSTHWGFVPVSE 450
Db 1254 NHEYLPVIRKIEHYASLLESTSTHWGFVPASE 1285

RESULT 3
Q45967
ID Q45967 PRELIMINARY; PRT: 1285 AA.
AC Q45967
DT 01-NOV-1996 (Tremblrel. 01, Created)
DE Neurotoxin consisting of botulinum neurotoxin D and C1.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE D;
MDLINE=96283801; PubMed=8679691;
RA Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA Ogumada K.;
RT "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT types C and D organisms";
RL Biochim. Biophys. Acta 1307:123-125(1996).
DR EMBL; D38442; BAA07477.1; -
DR HSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1285 AA; 147365 MW; EDE98E4EAC6413 CRC64;

Query Match 76.2%; Score 1823; DB 2; Length 1285;
Best Local Similarity 77.7%; Pred. No. 1.5e-87;
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

Qy 2 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGQVQLN 61
Db 839 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGQVQLN 898
Qy 62 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 121
Db 899 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 958
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYKWFVTVTNMGMNMI 181
Db 959 KNSGWSIGIISNLFVTLKQNESEQDINFSYDISKNAAGYNKWFVTVTNMGMNMI 1018

```

```

QY 182 YINGKLIDITIKVKELTGINFSTIIFKINPKIPDTGLITSDSDINNMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDITIKVKELTGINFSTIIFQNKIPNTGLITSDSDINNMWIRDFYIFAKELDD 1078
QY 242 KOINILFNSLQTNVVKDYWGNDLRVKNKEYYMWINDYLNRYMYANSROIVFNTRENNDF 301
Db 1079 KOINILFNSLQTNVVKDYWGNDLRVKNKEYYMWINDYLNRYMYANSROIVFNTRENNDF 1138
QY 302 NEGYKIIIRKIRGNTNDTRVRGDDILYFDMTINNKAYLNFMKNTMYADNHST-EDIYAI 360
Db 1139 NEGYKIIIRKIRGNTNDTRVRGDDILYFDMTINNKAYLNFMKNTMYADNHST-EDIYAI 1193
QY 361 GUREQTKD-INDNIIFOIPOMNNYVYASQIFKSNFNGENISGICSTGVYFRIGSD-WY 418
Db 1194 GALDOPMDEIRYGFIIOPCNTFYASQIFKSNFNGENISGICSTGVYFRIGSD-WY 1253
QY 419 RENVLPVTVKQGNYSALLESTSTHGWFPVSE 450
Db 1254 NHEYLIPVTKIEHYASALLESTSTHGWFPVSE 1285

RESULT 4
Q9LBS7 PRELIMINARY; PRO: 1280 AA.
AC Q9LBS7:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-6814;
RA Sagane Y., Watanabe T., Kouguichi H., Yamamoto T., Kawabe T.,
RA Murakami F., Nakatsuka M., Ohyama T.;
RT "Organization of Gene Encoding Components of the Botulinum Progenitor
RT Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT Chimeric Sequence in the Gene Encoding Each Component.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB037166; BAA099713.1;
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1280 AA; 147757 MW; DEL24FFB6F68450B CRC64;

Query Match
Best Local Similarity 36.9%; Score 882.5; DB 2; Length 1280;
Matches 198; Conservative 78; Mismatches 148; Indels 27; Gaps 10;

QY 2 TIPFNIFSYTNNSLLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 61
Db 843 TIPFNIFSYTNNSLLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 902
QY 62 PTFPDEKLGSGEDRGKVIYVTONENIYVNSMYESFSISFWIRNKWVSNLPG-YTIIDS 120
Db 903 TIYTNDFKLSSGD---KIIVNLNNILYSAIYENSSVFWIKSKDLTNSHNEYTIINS 959
QY 121 VKNNSGWSIGTISNLFVTLTKONEDSEOSINFSDISNNAPGY-NKWFVYVTNNMGMN 179
Db 960 IKQNSGWLKICRNGNIEWLQDINRKYSLFDYSESLSHGTGYNKFWFTITNNINGYM 1019
QY 180 KYINGKLIDITIKVKELGINFSTIIFKINPKIPDTGLITSDSDINNMWIRDFYIFAKEL 239
Db 1020 KLYINGELKQSERIEDLNEVKLDTIVFGDENID-----ENQMLWIRDFNIFSKEL 1071

```

```

QY 240 DGKDNILFNSLQTNVVKDYWGNDLRVKNKEYYMWINDYLNRYMYANSROIVFNTRENN 299
Db 1072 SNEDINIVIEGOILRNVIKDYWGNPLKEDFYIINNDYIDRYIAPKSNILVLVOYFDRS 1131
QY 300 DFNEGYKIIIRKIRGNTNDTRVRGDDILYFDMTINNKAYLNFMKNETMYA---DNHSTED 356
Db 1132 KLYTCNPITIKSVSDKNPYSRILGNDINMFHLYNSGKMIIRDTDIYATEGRECSKNC 1191
QY 357 IYAIGLRQTKDINDNIIFOIPOMNNYVYASQIFKSNFNGENISGICST-GTYRFRLLG 415
Db 1192 VYALQSNLNGYGG-IFSINIVSQNKYSQIFSSFM---KNTMLLADYKPPWRFSE-- 1246
QY 416 DWYRHNYLVPVTVKQGNYSALLESTSTHGWGFV 446
Db 1247 ---ENAVTP-VAVTNYETKLLSTSTFWKFI 1272

RESULT 5
Q45849 PRELIMINARY; PRT: 1280 AA.
AC Q45849;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Neurotoxin.
OS Clostridium botulinum C.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=36828;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6813;
RX MEDLINE-96156810; PubMed=8593068;
RA Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA Oguma K.;
RT "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT detection of this gene from Clostridium botulinum type C organisms.";
RL Appl. Environ. Microbiol. 62:662-667(1996).
DR EMBL; D49440; BAA08418.1;
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1280 AA; 147815 MW; 74F813B228B8C989 CRC64;

Query Match
Best Local Similarity 36.9%; Score 882.5; DB 2; Length 1280;
Matches 198; Conservative 78; Mismatches 148; Indels 27; Gaps 10;

QY 2 TIPFNIFSYTNNSLLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 61
Db 843 TIPFNIFSYTNNSLLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 902
QY 62 PTFPDEKLGSGEDRGKVIYVTONENIYVNSMYESFSISFWIRNKWVSNLPG-YTIIDS 120
Db 903 TIYTNDFKLSSGD---KIIVNLNNILYSAIYENSSVFWIKSKDLTNSHNEYTIINS 959
QY 121 VKNNSGWSIGTISNLFVTLTKONEDSEOSINFSDISNNAPGY-NKWFVYVTNNMGMN 179
Db 960 IKQNSGWLKICRNGNIEWLQDINRKYSLFDYSESLSHGTGYNKFWFTITNNINGYM 1019
QY 180 KYINGKLIDITIKVKELGINFSTIIFKINPKIPDTGLITSDSDINNMWIRDFYIFAKEL 239
Db 1020 KLYINGELKQSERIEDLNEVKLDTIVFGDENID-----ENQMLWIRDFNIFSKEL 1071
QY 240 DGKDNILFNSLQTNVVKDYWGNDLRVKNKEYYMWINDYLNRYMYANSROIVFNTRENN 299

```

```
Db 1072 SNEDINIVEGQILRNVIKDYWGNGPLKFDTEYYIINDNYIDRYIAPKSNILVLVQVDRS 1131
Qy 300 DFNEGYKIIIRKRGNTNDVRGSGDILYFDMTNNKAYNLFMKNETMYA---DNHSTED 356
Db 1132 KLYTGNPTITKSDKPNYSRLNGDNIMHMLYNSGKYMIRDTDTIYVAGRECSKNC 1191
Qy 357 IYATGLRQTKDINDNIIFOIQPMNNYYIASOIFKSNFNGENISGICSI-GTYFRLLGG 415
Db 1192 VYALKQLSNLNGYIG-IFSINKTVSONKYSQIFSSFM--KNTMLLADIYKWPRESF-- 1246
Qy 416 DWYRHNYLPTVKQGNVSLLESTSTHWGFV 446
Db 1247 ---ENAYTP-VATNVTETKLLSTSSFWKFI 1272

RESULT 6
Q90TG7 PRELIMINARY; PRT; 1275 AA.
AC Q90TG7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NTX (Fragment).
GN Clostridium botulinum D bacteriophage.
OC Viruses.
OX NCBI_TaxID=29342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1873;
RA Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT "Molecular composition of the 16S toxin produced by a Clostridium
RT botulinum type D strain, 1873."
RL Microbiol. Immunol. 42:599-605(1998).
DR EMBL; AB012112; BAA75084.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; Bontoxilysin.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1275 1275
SQ SEQUENCE 1275 AA; 146742 MW; 3C50F46C8233E2D6 CRC64;

Query Match 36.3%; Score 867.5; DB 12; Length 1275;
Best Local Similarity 43.2%; Pred. No. 1.3e-37;
Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

Qy 2 TIFPNIFSYTNNLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61
Db 839 TMPFNIFSYTNNLLKDIINEYFNINDSKLSLQNKALVDTSGYNAEVRVGDVQLN 898
Qy 62 PIFPFDKLGSGEDRGKIVTQENIVYNSMYESISFWIRKWNLPQ-YTIIDS 120
Db 899 TIYTNDFKLSSGD--KIIYVNLNLLYSAIYENSVSFWIKSKDLTNSHNEYTIIS 955
Qy 121 VKNNSGWSIGIISNLFVTLTKQNEDEQSINFSYDVSNNAPGY-NKWFVFTVNNMGNN 179
Db 956 IEQNSGKLCIRNGNIEWLODVRKYSLLFDYSSELSHTGYNKWFVFTVNNMGYK 1015
Qy 180 KIYNGKLIDIKVKELTGNFSKTIPEINKEIPDTGLTSDSDNNIMWIRYIPAKEL 239
Db 1016 KLYINGELKQSKIDLEDVKLDTIFVIGDENID-----ENQMLTDFNIFPSKEL 1067
Qy 240 DGKDNILFNSLQVTNVYKDWGNDLRYNKEYYMNIDYLNRYMYANSRQIVFTRNNN 299
Db 1066 SNEDINIVEGQILRNVIKDYWGNGPLKFDTEYYIINDNYIDRYIAPKSNILVLVQVDRS 1127
Qy 300 DFNEGYKIIIRKRGNTNDVRGSGDILYFDMTNNKAYNLFMKNETMYADNH---STD 356
```

```
Db 1128 KLYTGNPTITKSDKPNYSRLNGDNIMHMLYNSGKYMIRDTDTIYVAGRECSQNC 1187
Qy 357 IYATGLRQTKDINDNIIFOIQPMNNYYIASOIFKSNFNGENISGICSI-GTYFRLLGG 415
Db 1188 VYALKQLSNLNGYIG-IFSINKTVSONKYSQIF--SSFR-ENTMLLADIYKWPRESF--- 1241
Qy 416 DWYRHNYLPTVKQGNVSLLESTSTHWGFV 446
Db 1242 --FKNAVTPAVT--NYETKLLSTSSFWKFI 1268

RESULT 7
Q933KO PRELIMINARY; PRT; 1291 AA.
AC Q933KO;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=593, AND 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;

Query Match 25.5%; Score 610; DB 2; Length 1291;
Best Local Similarity 33.9%; Pred. No. 3.7e-24;
Matches 168; Conservative 84; Mismatches 154; Indels 90; Gaps 20;

Qy 2 TIFPNIFSYTNNLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61
Db 835 SIPFDLSTYNTNLTILIEFPKNYSIDLNLNRYRDNKLDLSYGAKEVYDGVKLN 894
Qy 62 PIFPFDKLGSGEDRGKIVTQENIVYNSMYESISFWIRKWN---VSNL--PGY 115
Db 895 D--KNQPKLTSSA--NSKIRVTQNTQNIIFNSMFLDPSVFWIRPKYKNDGIONTHNEY 950
Qy 116 TIIDSVKNNSGWSIGIISNLFVTLTKQNEDEQSINFSYDVSNNAPGY-NKWFVFTVNN 174
Db 951 TIINCNNNSGWSIRGNMIWTLIDINGKIKSVFEYSIKEDISEYINRWFVFTVNN 1010
Qy 175 MGNMKYIYNGKLIDTKVKELGINFPSKTIPEINKEIPDTGLTSDSDNNIN---MWIR 230
Db 1011 -SDNAKIYIYNGKLESHIDIRDIANDIENDEIFKLD-----GNIDRTOFIMWK 1057
Qy 231 DFYIFAKELGKDNILFNSLQVTNVYKDWGNDLRYNKEYYMNIDYLNRYMY---AN 286
Db 1058 YFSFNTLQSQNIIEYIKQSYSEVLKDFWGNPLMYNKEYYMFNAGNKSNIKLLKSSD 1117
Qy 287 SRQIVFTRNNN-----DFNEGKIIIRKRGNT---NDRVRGGDILYFDMTINNKK 336
Db 1118 VGEILLTRSKYQNQSKYINRYDRLYIGKFTIRR-KNSQSQINDIDIVRKEDYIYLDFFNLQ 1176
Qy 337 AYNLFMKNETMYADNHSTEDIYALGLEQTK---DINDNIIF---QIQPMNNIYYVAS 388
Db 1177 EWRVIM-----YKFKKEEKLFLAPISDSDEFYNTIQIKEYDEQPTVSC 1221
```



Qy	2	TIPFSTSTNNLSLLKDIINEFPNNINDSKILSLQNRKNTLVDTSGYNAEVSSEGDVQLN	61
Db	835	SIPFDLSTYNTNTILIEIFNQSNDILNINILNRYDRONKLIDISGYAKVYDGYVKLN	894
Qy	62	PIPPFQFKLSSGEDRGKIVYQENINIVNSMYESFSFWIRNKW---VSLN--PGY	115
Db	895	D--KNQFKLTSSA--NSKIRVIONQNIIFNSMFLDFSFWIRIPKYNKDGIQNIHNEY	950
Qy	116	TIDTSVKNNSGWSIGIISNFIYFTVLKQNSDSQSINFSYDISNAPGY-NKWFPVTVTNN	174
Db	951	TIINCNNNSGWRISIRGNMILWTIDINGKIKISYFFYSIKEDISEYNKWFVTTINN	1010
Qy	175	MQGNMKIYIINGKLIDIKYKVELTGINFSTIFFENKIPDTGLITSDSNIN---MWIR	230
Db	1011	-SDNAKIYINGKLESHIDIRDIREVIANDLFIKLD-----GNIDRTQFIWMK	1057
Qy	231	DFYIFAKELDKGDIINFLNSLOTVNVDYWGNDIRYKKEYYMNIDYLRMYTANS----	287
Db	1058	YFSIFNTELSQSNEIEYIKQYSEYLDGFWNPJMYKNEYTMAGNKNYSYIKLKXDSP	1117
Qy	288	-ROIIVNTRNN-----DFNEGKIIIRKIRGNT---NDTRVRGSDIYLFDMTINNK	336
Db	1118	VGEILPRSKYNQSKYINRYDLVIGEKFPIIR-KNSQSINDDIVKREDIYILDFENLQ	1176
Qy	337	AYNLFMKNETMYADNHSTEDITYAIGUREQTK---DINDNIIF---QIQPMNNTYYAS	388
Db	1177	EWRYV-----IYKFKKESEKFLAPISDSDEFYNTIQIKEDYSQPTISC	1221
Qy	389	Q-IFKSNFNGENISGICSTGYRFLGG-----DWYRHNYLVPTVRCQGNVAS	434
Db	1222	QLLFKK--DRESDEIGLGIHRFYESGIVFKEYKDYFCISKWY-----LAEVRKPYNS	1274
Qy	435	LLSESTTHGCFVPVSE	450
Db	1275	KL---GCNQWQIFPKDE	1287

[illegible]

DB 1210 ILGFRADTVASTWY-----YTMRDHTNSNGCFWNFI 1242

RESULT 12  
Q08077 PRELIMINARY; PRT: 1291 AA.

ID Q08077 AC Q08077  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-JUN-1996 (TREMELrel. 01, Last sequence update)  
DE 01-JUN-2001 (TREMELrel. 17, Last annotation update)  
GN Botulinum neurotoxin type B (EC 3.4.24.-) (BONT/B).  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TAXID=1491;  
RN [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN=EKLUND 17B ATCC25765;  
RA MEDLINE=94122659; PubMed=7764370;  
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;  
RT "Nucleotide sequence of the gene coding for non-proteolytic  
RT clostridium botulinum type B neurotoxin: comparison with other  
RT clostridial neurotoxins";  
RL Curr. Microbiol. 28:101-110(1994).  
CC -! FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
CC ENDOPEPTIDASE THAT CLEAVES SYNAPTOSOMAL VESICLE  
CC -! SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A A  
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATES CHANNEL  
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC -! SUBCELLULAR LOCATION: SECRETED.  
CC -! MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
CC -! SIMILARITY: HIGH WITH OTHER BOTULINUM NEUROTOXINS AND WITH TETANUS  
CC NEUROTOXIN.  
CC -! SIMILARITY: TO OTHER ZINC METALLOPROTEINASES IN THE ACTIVE SITE  
CC REGION.  
DR EMBL: X71343; CAAS50482.1; -  
DR HSSP: P10845; 3BTA.  
DR MEROPS: M27.002; -  
DR InterPro: IPR000395; Bontoxilysin.  
DR InterPro: IPR000130; Zn\_Mtpeptidse.  
DR Pfam: PF01742; Peptidase\_M27; 1.  
DR PRINTS: PR00760; BONTTOXILYSIN.  
DR ProDom: PD001963; Bontoxilysin; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 24.8%; Score 593; DB 2; Length 1291;  
Best Local Similarity 32.9%; Pred. No. 2,9e-23;  
Matches 162; Conservative 87; Mismatches 159; Indels 84; Gaps 20;

QY 3 IPNFIFSYTNNSLLKDIINEYFNPNINDSKILSNQRKNLTVDTSYNAEVSSEGVOLNP 62  
Db 836 IPPDLSTYSNIEIILIKFNKYNSEILNNIIILRYRDNRLIDLSDSGAKVEYDVGLND 895  
QY 63 IPFFDFKLSSGEDRGKVVTQNETNVNMYSEFSISFWIRNKW---VSNL--PGYT 116  
Db 896 --KNQPKLYSSAD--SKIRVTQNQLIENSMLDFSVSWIRPIRYRDNDDQLNYIHNEYT 951  
QY 117 IIDSVKNNNSGWSIGTISNFLFTLKQNEDESGSINFSDISNNAGY-NKWFVFVTTNMN 175  
Db 952 IINCMMNNNSGWKISIRGNRIIWLTDINGTKTSVFEYNIREDISYINRWFFVTTNN- 1010  
QY 176 MGMKKYINGKLDITTKVELTGINFSTITEINKIPDGTGLITSDSNIN-MWRDPYI 234  
Db 1011 LDNAKIYINGLESNMMDIKDIGEVINGBTIKFLD-----GDVRTQFIWKKYESI 1061

Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
"C. butyricum (KZ 1890) gene for type E botulinum toxin.";  
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[10]  
RN SEQUENCE FROM N.A.  
RC STRAIN=KZ 1891;  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
"C. butyricum (KZ 1891) gene for type E botulinum toxin.";  
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[11]  
RN SEQUENCE FROM N.A.  
RC STRAIN=LCL 063;  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
"C. butyricum (LCL 063) gene for type E botulinum toxin.";  
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN EMBL: AB037714; BAB03522.1; -  
DR EMBL: AB037704; BAB03512.1; -  
DR EMBL: AB037705; BAB03513.1; -  
DR EMBL: AB037706; BAB03514.1; -  
DR EMBL: AB037707; BAB03515.1; -  
DR EMBL: AB037708; BAB03516.1; -  
DR EMBL: AB037709; BAB03517.1; -  
DR EMBL: AB037710; BAB03518.1; -  
DR EMBL: AB037711; BAB03519.1; -  
DR EMBL: AB037712; BAB03520.1; -  
DR EMBL: AB037713; BAB03521.1; -  
DR HSSP: P10845; 3BTA.  
DR MEROPS: M27.002; -  
DR InterPro: IPR000395; Bontoxilysin.  
DR InterPro: IPR000130; Zn\_Mtpeptidse.  
DR Pfam: PF01742; Peptidase\_M27; 1.  
DR PRINTS: PR00760; BONTTOXILYSIN.  
DR ProDom: PD001963; Bontoxilysin; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 1251 AA; 143751 MW; 2021F4E427070296 CRC64;

Query Match 24.8%; Score 594; DB 2; Length 1251;  
Best Local Similarity 32.2%; Pred. No. 2,5e-23;  
Matches 150; Conservative 94; Mismatches 156; Indels 66; Gaps 16;

QY 2 TIPNFISYTNNSLLKDIINEYFNPNINDSKILSNQRKNLTVDTSYNAEVSSEGVOLNP 61  
Db 822 SIPFKLSYSTDKILISYFNKFKRIKSSVMNMYKNDKYVDTSYDSNININGEIFY 881  
QY 62 PIFFDFKLSSGEDRGKVVTQNETNVNMYSEFSISFWIRI---NKVSNLPGYTI 117  
Db 882 PTNKNQFTIFS--KSEVNISONDYIIIDNKKYKFSISFWIRIYDNKVINNETI 939  
QY 118 IDSVK--NNSGSIGTISNFLFTLKQNEDESGSINFSDISNNAGY-NKWFVFVTTNMN 175  
Db 940 INCMMNNNSGWKSLNHBNIIWTLDQNRARKLVKFYGNANGISDIYINKWIFVITI 999  
QY 176 MGMKKYINGKLDITTKVELTGINFSTITEINKIPDGTGLITSDSNIN-MWRDPYI 235  
Db 1000 LGPSKLYINGHLIDQRSILNLGNIRHVSNDILFKVNCSTRYI-----GIIFYNIF 1050  
QY 236 AKELDGDKDINLFNSLQYTNVVKYWDYNGNDRLNKYYVMYNDILNRYM--YANSRQIVFN 293  
Db 1051 DKELDETEOTLYSNPPNNILKDFNGVLLYDKGYLLNLVLPKPNFIDRRKOSTLSINN 1110  
QY 294 TRN---NNDNFEGYKIIKRIR-GMNIDTRVRGGDILYFDMTINKAYNIFMKWETHYA 349  
Db 1111 IRSTILLANLEYSGIKVIQRVNDSTNDRFVRKNDQVYINTVISNSSYSL-----YA 1163  
QY 350 DNSTEDIYAIUREQT--KDINDNIIFIQPMNTTYIYASQIFKSFNENGENISGICSIG 407  
Db 1164 DTWTD-----KEKTIKSSSGNRFNQVVVYNSVGNNCTNMFANN-NGNN-----IG 1209  
QY 408 TYRFR----LGSGDVRHNYLVPTVKOGNYASLLESTSH---WGVF 446







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: November 7, 2002, 17:45:16 ; Search time 1950 Seconds  
(without alignments)  
3737.420 Million cell updates/sec  
Title: US-09-910-186a-10  
Perfect score: 2391  
Sequence: 1 MTIFNFSTNNSLLKDI.....NYASLLESTHGWGFPVSE 450

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09910186/runat\_04112002\_111617\_20848/app\_query.fasta\_1.647  
-DB=EST -QEXT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORES=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-USER=US09910186 -CGN\_1\_1899 -runat\_04112002\_111617\_20848 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEVTIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_oth:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127.5	5.3	870	17	AZ675343 ENTIG64TF
2	122.5	5.1	929	17	BH167076 ENTQX64TR
3	122	5.1	3707	14	C83838 C83838 Dict
4	119.5	5.0	917	17	AZ684830 ENTHZ58TF
5	119	5.0	649	9	AT297914 LP12358.5
6	117	4.9	575	17	AA549915 0988m3 gm
7	116	4.9	562	13	BJ360994 BJ360994
8	116	4.8	641	12	BF297786 056PBC01
9	114.5	4.8	664	10	AW584247 N210227e
10	114	4.8	653	17	AA550238 1358m3 gm
11	113.5	4.7	846	17	AZ689135 ENTGX61TF
12	112	4.7	549	13	BI813874 PfESToa1
13	112	4.7	943	17	BH134547 ENTOL08TF
14	110	4.6	500	9	AU087833 AU087833
15	110	4.6	561	13	BM275413 PfESTOa6
16	110	4.6	690	13	BM162769 EST565303
17	110	4.6	945	17	AL404242 T7 end of
18	109.5	4.6	696	13	BM162617 EST565140
19	109.5	4.6	720	13	BM169137 EST571560
20	109.5	4.6	781	13	BM166487 EST569010
21	109	4.6	571	13	BM274132 PfESTOa6
22	108.5	4.5	597	14	BQ597280 PfESTOa3
23	108	4.5	638	14	BM815053 EST593147
24	108	4.5	754	17	AA550316 1457m3 gm
25	108	4.5	796	13	BM159196 EST561719
26	108	4.5	877	17	AZ683138 ENTML73TR
27	107.5	4.5	767	13	BM162128 EST564851
28	107	4.5	920	17	AZ547074 ENTEN03TF
29	107	4.5	568	13	BM275819 PfESTOa8
30	107	4.5	610	14	BQ595977 PfESTOa2
31	107	4.5	722	13	BM164970 EST567493
32	107	4.5	1225	9	ABO12854 ABO12854
33	106.5	4.5	662	17	AZ524207 228PBD12
34	106	4.4	561	13	BM274447 PfESTOa4
35	106	4.4	732	13	BM159156 EST561579
36	106	4.4	950	17	AZ693368 ENTJUL19TR
37	105.5	4.4	688	12	BG604058 EST503148
38	105.5	4.4	694	12	BG604057 EST503147
39	105.5	4.4	698	13	BM163799 EST566322
40	105.5	4.4	764	13	BJ430749 BJ430749
41	105	4.4	1110	17	CNS06PWT
42	104.5	4.4	622	17	AZ526922 T3 end of
43	104.5	4.4	687	17	AZ526437 261PBC05
44	104.5	4.4	750	13	BM164094 EST566617
45	104.5	4.4	812	13	BM163086 EST565609

ALIGNMENTS

RESULT 1  
AZ675343  
LOCUS  
DEFINITION  
ENTIG64TF Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000  
genomic, DNA sequence.  
ACCESSION  
AZ675343  
VERSION  
AZ675343.1 GI:11812609  
KEYWORDS  
GSS.  
SOURCE  
Entamoeba histolytica.  
ORGANISM  
Entamoeba histolytica.  
REFERENCE  
1 (bases 1 to 870)  
AUTHORS  
Loftus, B., Van Aken, S. and Fraser, C.  
TITLE  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library

JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 23  
High quality sequence stop: 724.

## FEATURES

Location/Qualifiers  
1..870  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds M. Vaudin and B.  
Bareil, Oxford University Press, 1999)."

BASE COUNT 376 a 112 c 126 g 256 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4.26e-05 Length: 870  
Score: 127.50 Matches: 71  
Percent Similarity: 37.58% Conservative: 44  
Best Local Similarity: 23.20% Mismatches: 116  
Query Match: 5.33% Indels: 75  
DB: 17 Gaps: 13

US-09-910-186a-10 (1-450) x A675343 (1-870)

QY 7 IlePheSerTyThrAsn-----AsnSerLeuLeuLys 17  
Db 18 ATCTTTACATTAACATCTATTTGCGATGATAACAAGATCAATATAACAACATTCAGACAA 77  
QY 18 AspIleAsnGluTyPheAsnAsnIleAsnAspSerLysIleLeuSer----- 34  
Db 78 GACATAATAATAATCACTAATGTTTCACCAATAAACAACAACACAGTGCAC 137  
QY 35 -----LeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyAsnAlaGluVal 52  
Db 138 CCACATAATAATGTAACATGTTTCCAAAATACCTCCAAAATACCTACAGATGTG 197  
QY 53 SerGluGluGlyAspValGlnLeuAsnProIlePhePropheAspPheLysLeuGlySer 72  
Db 198 TCIGAA-----CCGAATGTCCTAATGAAGGAGTATAAGATTG 236  
QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyAsnSer 92  
Db 237 GAAAGTCATGAAATGAAATGAAATGAAATATATATATATATATATATATATATATATAT 296  
QY 93 MetTyrgLysPheSerIleSerPheTyrPhe----- 103  
Db 297 AAGATGAGATGATGCAATGCTATCTAGTTGTGTCACACAGGAGTGAAGAACT 356  
QY 104 ---ArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerValLys 122  
Db 357 ACATTATTGAATAGTTTGTAAATGCAATATATATATATATATATATATATATATATATAT 416

QY 123 AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGln 142  
Db 417 TAT-----ATAATTATAAT-----GAAGATAATTATGAACAA 449  
QY 143 AsnGluAspSerGluGlnSerIleAsnPheSerTyAspIleSerAsnAlaProGly 162  
Db 450 CGTAAAGATCAATCTAAAGTCAACAAAGTCAA----- 482  
QY 163 TyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyr 182  
Db 483 -----GTAACAATTTATAAT-----ATTAAAGAACAAAAGAACAA 518  
QY 183 IleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSer 202  
Db 519 CCACCAATCAAAATAATACATACACCAAGATTTGGAGATACAGAGGAATTCCTATCAT 578  
QY 203 LysThrIleThrPheGluLeu-----AsnLysIleProAspThrGlyLeu 217  
Db 579 ATGGAATTAATCTAAACAATAAAAGAGCATTGAAATAAAGTATTAGATTAAATGCT 638  
QY 218 Ile-----ThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyIlePhe 235  
Db 639 ATTGTGTTTCTGCACAATCAAGTAATGTGAGGCTAACACTAAGTCAACACTATATATT 698  
QY 236 AlaLys-----GluLeuAspGlyLysIleAsnIleLeuPheAsnSerLeuGlnTyr 253  
Db 699 GGAATATTATCACTCACTATTATTTGGAAAGATGTAAAAGAAATT-----TAT 743  
QY 254 ThrAsnValValLysAspTyTrpGlyAsnAspLeuArgTyAsn-----Lys 269  
Db 744 TGCATGCTTACATTCGTGTGATGTTAAACACCAACAATTTATATGCTACATCAATA 803  
QY 270 GluTyTrpTyMetValAsn 275  
Db 804 GATGATTTTTCAGTCCAT 821

## RESULT 2

## BH167076/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

BH167076 929 bp DNA linear GSS 24-SEP-2001  
ENTOX64TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.  
BH167076  
BH167076.1 GI:15740514  
GSS.  
Entamoeba histolytica.  
Entamoeba histolytica.  
Eukaryota; Entamoebidae; Entamoeba.  
1 (bases 1 to 929)  
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library (2001)  
Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 17  
High quality sequence stop: 592.  
Location/Qualifiers  
1..929  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The

FEATURES  
source



```
Qy 109 ValSerAsnLeuProGlyTyrThrIleAspSerValLysAsnAsnSerGlyTrpSer 128
Db 1880 CAAGATGACCTATTGGT-----AGTAATAGTAGTAATCA 1915
Qy 129 IleGlyIleIleSerAsnPhLeuValPheThrLeuLysGln-----142
Db 1916 AAGATTATCAAAATCAACTCAATGTTTGCCTCATCAAGTCGATATCTGGTGGTAAT 1975
Qy 143 ---AsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
Db 1976 TGTAATAGTAGTAATCAAGATTAATCAAAATTAACAACTCAATGTTTGCCTCATCAAGT 2035
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 2036 GCATAT-----ACTGGTGGTAATTTTAACTAATACCATATC 2071
Qy 182 TyrIleAsnGlyLysLeuIleAspThr---IleLysValLysGluLeuThrGlyIleAsn 200
Db 2072 TATGCCAAATAAAGTATGATACCGGTGTAATTAACACCTTTCAATGGGCTCA 2131
Qy 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 2132 GCAAAACAATACCATATGATGTCAAATAAAGTTAATGATACCGGTGTAATATCA 2191
Qy 221 -----AspSerAspAsnIle-----225
Db 2192 ACCTTTTCAATGGGCTCAGCAGATGTTTATGTCGCAAGACAATTTTAACTTTAT 2251
Qy 226 -----AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLys 242
Db 2252 CAATTAATAATGAATGGGTACTGGTTCATATTTTAACTTTTATGTTTATGTCGA 2311
Qy 243 AspIleAsnIleLeuPheAsnSerLeuGluTyrThrAsn-----255
Db 2312 CAA-----TTTTTAACCTTTATCAATTAATGAATGAGGTGTTACTGTTCCAT 2362
Qy 256 ValValLysAsp-----TyrTrp-----261
Db 2363 ATTTTAAAAATAAATTTATTTACTAGTAAAAAGTGGTATCACTCAAAATGAATGCCCA 2422
Qy 262 GlyAsnAspLeuArgTyr-----AsnLysGluTyrTyr-----272
Db 2423 GTATGATCAAGATGCTTTATTAAACAATAAATTTATTACTTACTAGTAAGGTGATC 2482
Qy 273 -----MetValAsnIleAspTyrLeuAsn 280
Db 2483 ACTCAATGAATGCCAGTACTGATCAAGATGCTTTATTAAACAATGAATGTCGAA 2542
Qy 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsn-----297
Db 2543 TCATTTGAATTTGTAATTAATGAATAACAATTTATTATTATACACAATGGTAAAT 2602
Qy 298 AsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgGlyAsnThrAsn 317
Db 2603 AAGAAGAATCTGCGAATCAATTTGAATTTGA-----AATAATGAA 2644
Qy 318 AspThrArgValArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAla 337
Db 2645 AAAACA-----ATTATTAT---ATTACAACAATGTTAAAT 2680
Qy 338 TyrAsnLeuPheMetLysAsn-----GluThrMetTyrAlaAspAsnHisSerThr 354
Db 2681 AAGAAATCGATGTTGAAATTTAGTTTCATCAACACTTTAT---GATTCACTACTGAT 2737
Qy 355 GluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIle 374
Db 2738 AAA-----TCATTAGGTGTTGTAATGTCACCCCGATGTT---GAAATTTAGTT 2785
Qy 375 PheGlnIleGlnProMetAsnAsnThrTyrTyr-----385
Db 2786 -----TCATCAACACTTTATGATTCATCTACTCTGATAATCATTAGGT 2827
```

```
Qy 386 -----TyrAlaSerGlnIlePheLysSerAsn 394
Db 2828 GTTTGTGAATGTGCACCAGGTTTCTCTGGAATTTATGTAATCAATGTTTAAATGTTACA 2887
Qy 395 PheAsnGlyGluAsn-----IleSerGlyIle-----403
Db 2888 GTTCAATGGGAATGTAATCAACATGTTTAAAGTTTCTCTGGAATTTATGTTAAT 2947
Qy 404 ---CysSerIleGlyThrTyrArgPheArgLeuGly 414
Db 2948 CAATGTGTTAATGTACAGTTCAATGGGAAATGGT 2983

RESULT 4
LOCUS AZ684830 917 bp DNA linear GSS 14-DEC-2000
DEFINITION ENT02581F Entamoeba histolytica Sheared DNA Entamoeba histolytica
geomic, DNA sequence.
ACCESSION AZ684830
VERSION AZ684830.1 GI:11821976
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE Eukaryota; Entamoebidae; Entamoeba.
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b1loftus@tigr.org
Clones are derived from the Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 24
High quality sequence stop: 871.
FEATURES
source
1..917
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT 390 a 104 c 153 g 270 t

ORIGIN

Alignment Scores:

Pred. No.: 0.000489 Length: 917  
Score: 119.50 Matches: 77  
Percent Similarity: 34.59% Conservatives: 61  
Best Local Similarity: 19.30% Mismatches: 116  
Query Match: 5.00% Indels: 145  
DB: 17 Gaps: 17

US-09-910-186A-10 (1-450) x AZ684830 (1-917)



```

...      |||...      |||...      |||...      |||...
Db 168 ACGTTCATCACTCAACAAATAAATGGATGTGTGTCATATTCACGACCGGATAT 227
Qy 218 -----IleThrSerAspSerAspAsnIleAsnMetTrp-----IleArgAspPheTyr 233
Db 228 TACGCTGCACCTACTCCGACGACAT-----TGCATGCCATCCACACCTCTG 278
Qy 234 IlePheIleAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253
Db 279 ATCAGACCCAAAC-----TGGGGAGGAATTCACGAG 308
Qy 254 ThrAsn-----ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGlu 270
Db 309 AACATCGCCGCCAGATCGTGGATGATCTGTC--AATCTGCCAGGCTGGATGTT 365
Qy 271 TyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIle 290
Db 366 ACCTACACACCTTACTCTGGATGTGATC--GAGTATTTCGACGCGGAGACCACTACATC 422
Qy 291 ValPheAsnThrArgAlaGAsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310
Db 423 CCTGGACATCGGCTTCAC-----GGCTTCACACTTGACTATC 464
Qy 311 ArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAsp 330
Db 465 GCGCTC-----GGAAACGACACG-----GCCGATTTCATTTACTATATC 503
Qy 331 MetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlnThrMetTyrAlaAsp 350
Db 504 CAGACCTTGACCAACAGGCTACATCAGCTGGCTTCATGAA-----GCATCG 554
Qy 351 AsnHisSerThrGluAspIleTyr 358
Db 555 AATGATACAGCTTGGACATTTC 578

RESULT 6
AA549915
LOCUS
DEFINITION
  0988m3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic
  clone 0988m, DNA sequence.
ACCESSION
  AA549915
VERSION
  AA549915.1 GI:2320167
KEYWORDS
  GSS.
SOURCE
  malaria parasite P. falciparum.
  Plasmodium falciparum
  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
  1 (bases 1 to 575)
AUTHORS
  Dame, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,
  Coppel, R., Cowman, A., Craig, A., Fischer, K., Foster, J., Goodman, N.,
  Hinterberg, K., Holder, A.A., Holt, D., Kemp, D., Lanzer, M., Lim, A.,
  Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su
  X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, S.M.
  Current status of the Plasmodium falciparum genome project.
  Mol. Biochem. Parasitol. 79, 1-12 (1996)
97001675
CONTACT: Dame JB
  Dept. of Pathobiology, College of Veterinary Medicine
  University of Florida
  2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
  Tel: 352 392 4700
  Fax: 352 392 9704
  Email: damej@mail.vetmed.ufl.edu
  Seq primer: T3
  Class: shotgun.
  Location/Qualifiers
    1..575
      /organism="Plasmodium falciparum"
      /db_xref="taxon:5833"
      /clone="0988m"
      /lab_host="E. coli XL1-Blue"
      /note="vector: pBluescript SK(+); Genomic DNA, from
      asynchronous blood stage parasites of the cloned Honduran

```

HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 500C (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were ligated using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBluescript SK(+). Recombinant plasmids transformed E. coli XL1-Blue.

BASE COUNT 286 a 45 c 68 g 175 t 1 others  
ORIGIN

Alignment Scores: 0.000455 Length: 575  
Pred. No.: 117.00 Matches: 49  
Score: 36.73% Conservative: 23  
Percent Similarity: 25.00% Mismatches: 76  
Best Local Similarity: 4.89% Indels: 48  
Query Match: 17 Gaps: 8  
DB:

US-09-910-186a-10 (1-450) x AA549915 (1-575)

```

Qy 173 AsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly----- 185
Db 48 AATAATATTATGGAAGAAAGAAAGAAATATTTAAATGCTGTTCTAAAGAACAAACAGAA 107
Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
Db 108 AATTGCGATGATATATATCCAAAAATGAATATATATGATGATAAACAATAAACAACA 167
Qy 206 ThrPheGluIleAsn---LysIle-----ProAspThr 215
Db 168 CAATATTTTAAATAATGAAATTTTATTACATTTTAAATATATATGCAAAACAGAT 227
Qy 216 GlyLeuIleThrSerAsp-----SerAspAsnIleAsnMetTrpIleArgAspPheTyr 233
Db 228 TCATTAGATATAGATGATTTTAAATGATGAAAAAATATGACCTGTTTGAATGTTA 287
Qy 234 IlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253
Db 288 ATTAATAGAAACAACTTTGATAAAATGATTGGAATATTTTAAATATGTTACAAATTA 347
Qy 254 ThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMet 273
Db 348 AAAAATAATGAAAGAGAAAGAAATGAAAAAGAAATATTAATCATATGAATATTTCAAG 407
Qy 274 ValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsn 293
Db 408 GAAATATATACCAT---GATAAATAGTATATATACAAATCAATCAATATGCTTAAT 464
Qy 294 ThrArgArgAsnAsn---AsnAspPheAsnGluGlyTyrLysIleIleLysArgIle 312
Db 465 ATGAATTATACACTCCCAATAGGTACACGAGGAGTAATAATATT----- 509
Qy 313 ArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyrPheAspMetThr 332
Db 510 -----AATACTAATCT----- 521
Qy 333 IleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlnThrMetTyr 348
Db 522 -----AATGTTGNTCAAGGAATGAAACAACTAT 551

```

RESULT 7  
BJ360994  
LOCUS  
DEFINITION  
 BJ360994 Dictyostelium discoideum cDNA library, CF Dictyostelium  
 discoideum cDNA clone ddc8p21 5', mRNA sequence.  
ACCESSION  
 BJ360994  
VERSION  
 BJ360994.1 GI:19260589  
KEYWORDS  
 EST.  
SOURCE  
 Dictyostelium discoideum.  
ORGANISM  
 Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
REFERENCE  
 1 (bases 1 to 562)

/lab\_host="Wistar rats"  
 /note="Vector: pBluescript II vector DNA, excised from  
 lambda Zap II.; Site1: EcoRI; Site2: XhoI; Total RNA  
 extracted from asynchronous blood stage forms of the  
 cloned ANKA isolate of *P. berghei*, grown in Wistar rat  
 30% parasitemia and 2-5% gametocytemia. Contaminating  
 white cells had previously been removed and final host  
 cell contamination estimated to be approximately 5%.  
 PolyA+ RNA was extracted and reverse transcribed using

```

0190 gta-ahol primer (Lambda har-11 cDNA cloning kit)
Stragene). Second round cDNA was made following the
manufacturer's protocol. EcoRI adaptors were ligated
the cDNA, and fragments were ligated into EcoRI/XhoI
digested vector."
BASE COUNT      332 a      68 c      85 g      156 t
ORIGIN
Alignment Scores:
Pred. No.:      0.000988      Length:      641
Score:          115.00      Matches:      62
Percent Similarity: 35.23%      Conservative: 31
Best Local Similarity: 23.48%      Mismatches: 83
Query Match:      4.81%      Indels:      88
DB:              12      Gaps:      14
US-09-910-186A-10 (1-450) x BF297786 (1-641)
Qy  139 ThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheserTyrAspIleSerAsn 158
    ||| ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db  59 ACTAATAAAAAGAAAGAAACACGAAACGGAACGAGATTCAC-----AAAAAC 109
Qy  159 AsnAlaProGlyTyrAsnLysTsrPhePheValThrValThrAsnAsnMetMetGlyAsn 178
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  110 AATGAAATAACAGGAAATAAGTGG-----GGTGGT 139
Qy  179 MetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGly 198
    ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db  140 TTGAGA----- 145
Qy  199 IleAsnPheserLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIle 218
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

```

[illegible]



Qy 254 ThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMet 273  
 Db 281 AACACGAAATAAATAATGGC---AATAATACACTGAATTTAACAAATCTAATATAAT 337  
 Qy 274 ValAsnLysAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsn 293  
 Db 338 AATAATAGATATAATTCAGAAATCAAAATATAATTTTACAGAAATTTTAAATAATAC 397  
 Qy 294 ThrArgArgAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArg 313  
 Db 398 TACATAAAATAGAAACAATTTTAAATCAATGAAATAC----- 433  
 Qy 314 GlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIle 333  
 Db 434 GGGATACAAATACACAAACCGGATAGGA-----GATGAAACGAAC 475  
 Qy 334 AsnAsnLysAlaTyrAsn---LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis 352  
 Db 476 AATAATAATAGTATAACATATTGTCATGACAAATCTAAGCAACTATGCAATAGT--- 532  
 Qy 353 SerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn 372  
 Db 533 ----- 547  
 Qy 373 IleIlePheGlnIleGlnPrometAsnAsnThrTyrTyrAlaSerGlnIlePheLys 392  
 Db 548 -----AATCCAAAATGGGAAATACACAGCATTAACAACATTTGCCTGAT 592  
 Qy 393 SerAsnPheAsn 396  
 Db 593 GGCAATTTCAAT 604

RESULT 9  
 LOCUS AW584247 664 bp mRNA linear EST 07-SEP-2000  
 DEFINITION N210227e MHAM Medicago truncatula/Glomus versiforme mixed EST  
 library cDNA clone MHAM-IF21, mRNA sequence.

ACCESSION AW584247  
 VERSION AW584247.1 GI:7261301  
 KEYWORDS Medicago truncatula/Glomus versiforme mixed EST library.  
 SOURCE Medicago truncatula/Glomus versiforme mixed EST library  
 ORGANISM Eukaryota; mixed EST libraries.

REFERENCE 1 (bases 1 to 664)  
 AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,  
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 ESTs from roots of Medicago truncatula after colonization with  
 Glomus versiforme  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Harrison M. J.  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73401  
 Tel.: 580-223-5810  
 Fax: 580-221-7380  
 Email: mjharrison@noble.org  
 Other name: MHAM-1c-Cl1; Date: 3/14/00; Updated to the Database of  
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is  
 available at 'http://chrysis.tamu.edu/medicago'.  
 Seq primer: T3.  
 Location/Qualifiers  
 1..664  
 /organism="Medicago truncatula/Glomus versiforme mixed EST  
 library"  
 /cultivar="Medicago truncatula genotype A17"  
 /db\_xref="taxon:119092"  
 /clone="MHAM-IF21"  
 /clone\_lib="MHAM"  
 /tissue\_type="roots colonized with Glomus versiforme"  
 /dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The library was  
 made from a mixture of RNA from each of these stages."

FEATURES  
 source

/lab\_host="E. coli strain XL0LR"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The cDNA was  
 directionally ligated into the Unizap XR vector from  
 Stratagene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-Zap phage using Ex-assist  
 helper phage and propagated in XL0LR cells."  
 BASE COUNT 297 a 142 c 96 g 129 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00122 Length: 664  
 Score: 114.50 Matches: 73  
 Percent Similarity: 31.61% Conservatives: 37  
 Best Local Similarity: 20.98% Mismatches: 95  
 Query Match: 4.79% Indels: 143  
 DB: 10 Gaps: 19

US-09-910-186A-10 (1-450) x AW584247 (1-664)

Qy 4 ProPheAsnIlePhe---SerTyrThrAsnAsnSerLeuLeuLysAspIle----- 19  
 Db 4 CCAGCCACACACTTACCAATCATATGAAAAACTTCGTACACGATCAATCAACAAC 63  
 Qy 20 --IleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArg 38  
 Db 64 AAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 102  
 Qy 39 LysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVal 58  
 Db 103 -----AACAGTACCCTGAGTAAAGTTAGTGACACA 132  
 Qy 59 GlnLeuAsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGly 78  
 Db 133 AAATACACACACA-----TCTTACAGCAGCAGCAGCAACAAT----- 171  
 Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98  
 Db 172 -----AACTACGACACAAACAGTATCTTCTAGTCTGATGACTTTAGC 216  
 Qy 99 IleSerPheThrIleArgIleAsnLysTyrPheValSerAsnLeuPro-----GlyTyrThr 116  
 Db 217 -----ACAAA-----AAGTTCCGACGACGACGATATAAC 246  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 247 TCGAACGAAAAATCAGAACAAACAACAC-----AACGATAACAAG 285  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 286 TATTCTTACAAACACAAACACAGGATCTCTAGCAACACAAAGTTTACAGAANA--- 342  
 Qy 157 SerAsnAlaProGlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMet 176  
 Db 343 -----GGATACAAC-----TCCATGGAACCCGGAACAAC 372  
 Qy 177 GlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeu 196  
 Db 373 AACACACACAAAGTATTCTTACAAACAACAACAACAAGAT----- 411  
 Qy 197 ThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGly 216  
 Db 412 -----TCTTCTAGCAACA-----AAGTTCCAGAGAAGGT 444  
 Qy 217 LeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAla 236  
 Db 444 ----- 444  
 Qy 237 LysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256  
 Db ----- 256

```

Db 445 -----TACAACTCAATGCAAAATCAGAAC--- 468
QY 257 ValysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle 276
Db 469 -----ACAACTAC----- 480
QY 277 AspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArg--- 295
Db 481 -----GAGAAGTATAACTACAAAC-----AAAGTTGCTGTTAATGACAAATAC 525
QY 296 -----ArgAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArg 313
Db 526 AGCTTCAAAAGTAAACAACTACAAATGAGTGGAG-----AGG 564
QY 314 GlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIle 333
Db 565 CAAGGAATGAGTCACACGAGAGTTATGGAAGTGGGAATAATTTTATGATGTTAACTCT 624
QY 334 AsnAsnLysAlaTyrAsnLeuPhe 341
Db 625 GAGGAGAAATATACCCACATTC 648

RESULT 10
AA550238/c 653 bp DNA linear GSS 05-DEC-2000
LOCUS 1358m3 gmbpfH3.1, G. Roman Reddy Plasmodium falciparum genomic
DEFINITION clone 1358m, DNA sequence.
ACCESSION AA550238
VERSION AA550238.1 GI:2320490
KEYWORDS GSS.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 653)
AUTHORS Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z.,
Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,J., Goodman,N.,
Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A.,
Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su
,X.-Z., Thompson,J.K., Vital,F., Wellens,T.E. and Werner,E.
Current status of the Plasmodium falciparum genome project
Mol. Biochem. Parasitol. 79, 1-12 (1996)
97001675
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: Shotgun.

FEATURES
Source
1. .653
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="1358m"
/lab_host="E. coli XL1-Blue"
/note="Vector: pBluescript SK(+); Genomic DNA, from
asynchronous blood stage parasites of the cloned Honduran
HB3 isolate cultured in vitro, was digested with HincII
nuclease in the presence of 30% formamide at 50oC (Vernick
, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic
Acids Research 16:5883-5896). The ends of the fragments
were ligated using T4 DNA polymerase, and the fragments
were ligated to EcoRV-cleaved and dephosphorylated
pBluescript SK(+). Recombinant plasmids transformed E.
Coli XL1-Blue."
193 a 84 c 50 g 326 t
BASE COUNT
ORIGIN
Alignment Scores: 0.00137 Length: 653
Pred. No.:

```

```

Score: 114.00 Matches: 49
Percent Similarity: 39.71% Conservative: 32
Best Local Similarity: 24.02% Mismatches: 81
Query Match: 4.77% Indels: 42
DB: 17 Gaps: 10

US-09-910-186a-10 (1-450) x AA550238 (1-653)
QY 253 TyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyr 272
Db 645 TATAATGACAAATAAG-----AAAAACAAAAAATAATAGTTATTATGAA 598
QY 273 MetValAsnIleAsnTyrLeuAsnArgTyrMetTyrAla----- 285
Db 597 ATCAATAATAATACATATATGAATGAACATGATATACAGATATTGAAATGACGTTTA 538
QY 286 AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyr 305
Db 537 AATAAAAAAATAACGATTGAATGAGAGCTCGTACCAATACATGGATGAT----- 484
QY 306 LysIleIleLysArgIleArgGlyAsnThrAsnAspThr----- 319
Db 483 -----ATTATTGTAAGTCATCATGGAATGTTATGATAAATAATATACTAGTAAACAT 430
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
Db 429 AATAGAGAAAGATCATATATAATGAGATGAAAGAAACACAGATATAATA----- 379
QY 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAla 359
Db 378 -----AAAAAAACACGTTATTGTTAGATGA-----AAAGATATGGAAGGA 337
QY 360 IleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro 379
Db 336 ATAGGAAAGAAAGAAAGAAATAAATAATGATAATATATATTTTAT----- 289
QY 380 MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGlu--- 398
Db 288 ---AATACTCATATAGTATATAATAATAGTCTCTATAGTATATAATAATGATATA 232
QY 399 ---AsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrp 417
Db 231 TATAGTGTGTGATATATGACATCTGTTATATATACAAATATGATATCAGGAGTCCCAAGC 172
QY 418 TyrArgHis-----AsnTyrLeuValProThrValLysGlnGly-----AsnTyr 432
Db 171 TATGCTCATCTACTTATTATTAATAACCAAGTAAATGATATATATCAAGGCTACCTAATTAT 112
QY 433 AlaSerLeuLeu 436
Db 111 AATAACATGATG 100

RESULT 11
AA550238/c 846 bp DNA linear GSS 14-DEC-2000
LOCUS ENTGX611F Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ689135
VERSION AZ689135
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 846)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208

```



RESULT 13  
BH134547/c  
LOCUS



obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: 40UP from Gibco  
High quality sequence stop: 424.

# FEATURES

## source

Location/Qualifiers  
1..561  
/organism="Plasmodium falciparum 3D7"  
/db\_xref="taxon:36329"  
/clone\_lib="Plasmodium falciparum 3D7 gametocyte cDNA library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from gametocyte poly(A)+ RNA by oligo d(T) priming, size-selected and directionally cloned into the EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR lambda vector (Stratagene). The primary library was mass excised as phagemid using the ExAssist helper phage (Stratagene). Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells. Clone availability: David Sibley, Washington University."

BASE COUNT 270 a 32 c 62 g 197 t

## ORIGIN

### Alignment Scores:

Pred. No.: 0.00341 Length: 561  
Score: 110.00 Matches: 56  
Percent Similarity: 39.67% Conservative: 40  
Best Local Similarity: 23.14% Mismatches: 74  
Query Match: 4.60% Indels: 72  
DB: 13 Gaps: 13

US-09-910-186a-10 (1-450) x BM275413 (1-561)

QY 176 MetGlyAsnMetLysIleThrIleAsnGlyLysLeu-----IleAspThr 190  
Db 4 TTAAGCAATATATATTATATAAATAATAAATCAAGTTAAATCCCATACCTAGAAAT 63  
QY 191 IleLysValIleGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsn 210  
Db 64 GTGTCATAAAGAACATTTTGTGTAACTAGTATTAATAATA----- 108  
QY 211 LysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArg 230  
Db 109 -----TTATACATTATGATGAAGAAATATAATAAT----- 141  
QY 231 AspPheThrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe----- 248  
Db 142 -----AACTAATAAAATAGATATGATTAAGGTTACATAAACAVATCATATGGTAAT 195  
QY 249 AsnSerLeuGlnIleThrAsnValIleLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268  
Db 196 AATAACAATAATTAATAATAATATATGTTTATTAAACAGAGATTATTATTGCTAT 255  
QY 269 LysGluTyrIleMetValAsnIleAsp-----TyrLeuAsnArgTyrMetTyrAla 285  
Db 256 GAATTTATGCAATTAATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 315  
QY 286 AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsn---GluGly 304  
Db 316 AATAAAT-----ATTGTCATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 366  
QY 305 TyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGly 324  
Db 367 AATAAACCTTTTTCATAATATGATTAACAAATGAGAT----- 405  
QY 325 AspIleLeuTyrPheAspMetThr---IleAsn---AsnLysAlaTyrAsnLeuPheMet 342  
Db 406 ---GTATGCAATATGATATTTCAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 462

QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
Db 463 AAAAGAGGAAATATG----- 477  
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
Db 478 -----AAATACAAAAATATA-----AAT 495  
QY 383 ThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
Db 496 ACATATATATATGGAATAATATATTT-----TTTGAATGAATATATATATATGAT 546  
QY 403 IleCys 404  
Db 547 ATATGT 552

Search completed: November 7, 2002, 19:10:18  
Job time : 1959 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 11:55:50 ; Search time 2773 Seconds  
(without alignments)  
14388.729 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371  
Sequence: 1 gaattcagatgaccatccc.....tctccgagtaataaggaaattc 1371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	GenEmbl :
1: gb.ba.*	AC115681 Dictyoste
2: gb.htg.*	AC096320 Rattus no
3: gb.in.*	AC115964 Dictyoste
4: gb.om.*	AC113774 Rattus no
5: gb.ov.*	AC104634 Homo sapi
6: gb.pat.*	AC094507 Rattus no
7: gb.ph.*	AP001046 Homo sapi
8: gb.pl.*	AP001751 Homo sapi
9: gb.pr.*	AC119697 Rattus no
10: gb.ro.*	AC119558 Rattus no
11: gb.sts.*	AC126213 Rattus no
12: gb.sy.*	AF251281 Synthetic
13: gb.un.*	AF153362 Dictyoste
14: gb.vi.*	A58946 Sequence 6
15: em.ba.*	AC115681 Dictyoste
16: em.fun.*	AC079176 Homo sapi
17: em.hum.*	AC126871 Rattus no
18: em.in.*	AC096493 Rattus no
19: em.mu.*	AC091616 Rattus no
20: em.om.*	AC117070 Dictyoste
21: em.or.*	AC016445 Drosophil
22: em.ov.*	AC105625 Rattus no
23: em.pat.*	AR097042 Sequence
24: em.ph.*	AR097041 Sequence
25: em.pl.*	
26: em.ro.*	
27: em.sts.*	
28: em.un.*	
29: em.vi.*	
30: em.htg_hum.*	
31: em.htg_inv.*	
32: em.htg_other.*	
33: em.htg_mus.*	
34: em.htg_pln.*	
35: em.htg_rtd.*	
36: em.htg_mam.*	
37: em.htg_vrt.*	
38: em.sy.*	
39: em.htgo_hum.*	
40: em.htgo_mus.*	
41: em.htgo_other.*	

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	665.2	48.5	4479	7	CBCPHGC1	X53751 Clostridium
2	665.2	48.5	4592	7	CSTC1TOX	D90210 Bacterioph
3	665.2	48.5	4712	1	CBPHNC1M	X71126 C.b.culm
4	665.2	48.5	9613	1	CBCTOX	X62389 Botulinum b
5	665.2	48.5	9689	1	CBP1NCTC1	X66433 Clostridium
6	665.2	48.5	11747	1	AB061780	AB061780 Clostridi
7	665.2	48.5	12297	1	CBCTONT	X72793 Clostridium
8	449.8	32.8	11631	1	AB037920	AB037920 Clostridi
9	448.2	32.7	4049	1	CLQTNCO	D38442 Clostridium
10	189	13.8	151705	2	AC117900	AC117900 Rattus no
11	188.4	13.7	176351	2	AC117007	AC117007 Rattus no
12	183	13.3	169163	2	AC115666	AC115666 Rattus no
13	180.2	13.1	163034	2	AC099432	AC099432 Rattus no
14	170.2	12.4	165337	2	AC096212	AC096212 Rattus no
15	169.4	12.4	1402	6	AR000030	AR000030 Sequence
16	169.4	12.4	1402	6	AR169141	AR169141 Sequence
17	169.4	12.4	1402	6	AX036246	AX036246 Sequence
18	168.8	12.3	1330	6	AR000029	AR000029 Sequence
19	168.8	12.3	1330	6	AR169140	AR169140 Sequence
20	168.8	12.3	1330	6	AX036243	AX036243 Sequence
21	168.8	12.3	1338	12	XXU22962	U22962 Synthetic b
22	168.8	12.3	97683	2	AC116548	AC116548 Dictyoste
23	165.6	12.1	236542	2	AC096320	AC096320 Rattus no
24	164.6	12.0	256774	2	AC116964	AC116964 Dictyoste
25	160.2	11.7	149228	2	AC113774	AC113774 Rattus no
26	158.8	11.6	115758	9	AC104634	AC104634 Homo sapi
27	157	11.5	108476	2	AC094507	AC094507 Rattus no
28	156.8	11.4	158063	9	AP001046	AP001046 Homo sapi
29	156.8	11.4	340000	9	AP001751	AP001751 Homo sapi
30	155.4	11.3	173540	2	AC119697	AC119697 Rattus no
31	155.2	11.3	131346	2	AC119558	AC119558 Rattus no
32	154.6	11.3	169966	2	AC126213	AC126213 Rattus no
33	153.2	11.2	1299	12	AF251281	AF251281 Synthetic
34	153	11.2	7921	3	AF153362	AF153362 Dictyoste
35	151.6	11.1	1313	6	A58946	A58946 Sequence 6
36	151.6	11.1	39369	2	AC115681	AC115681 Dictyoste
37	150.6	11.0	186278	9	AC079176	AC079176 Homo sapi
38	147.6	10.8	229502	2	AC126871	AC126871 Rattus no
39	145.8	10.6	108177	2	AC096493	AC096493 Rattus no
40	145.8	10.6	125026	10	AC091616	AC091616 Rattus no
41	144.2	10.5	156533	2	AC117070	AC117070 Dictyoste
42	143.8	10.5	157141	3	AC016445	AC016445 Drosophil
43	136.6	10.0	147671	2	AC105625	AC105625 Rattus no
44	136	9.9	5163	6	AR097042	AR097042 Sequence
45	136	9.9	5318	6	AR097041	AR097041 Sequence

ALIGNMENTS

RESULT 1  
CBCPHGC1  
LOCUS Clostridium botulinum C phase gene for C1 neurotoxin.  
DEFINITION X53751  
ACCESSION X53751  
VERSION X53751.1 GI:14905  
KEYWORDS C1/neurotoxin; neurotoxin; toxin.  
SOURCE Clostridium botulinum C phase.  
ORGANISM Clostridium botulinum C phase  
Viruses.  
REFERENCE 1 (bases 1 to 4479)  
AUTHORS Popoff M.R.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1990) Popoff M.R., Institut Pasteur, Anaerobies,  
25 rue du Docteur Roux, 75724 Paris Cedex 15, France

[illegible]



Db 4060 CATTGGGGTTTGTACCTGTAAGTGAATAA 4089

## RESULT 2

CSTCITOX 4592 bp DNA linear PHG 23-JUN-1999  
LOCUS Bacteriophage c-st (from C. botulinum) C1-tox gene for botulinum C1  
DEFINITION neurotoxin.

ACCESSION D90210.1 GI:217780  
VERSION botulinum toxin; neurotoxin.  
KEYWORDS Bacteriophage c-st (from C. botulinum type C-Stockholm) DNA.  
SOURCE Bacteriophage c-st  
ORGANISM Viruses.

REFERENCE 1 (bases 1 to 4592)  
AUTHORS Kimura.K., Fujii.N., Tsuzuki.K., Murakami.T., Indoh.T., Yokosawa.N.  
Yokosawa.N., Takeshi.K., Syuto.B. and Oguma.K.  
TITLE The complete nucleotide sequence of the gene coding for botulinum  
type C1 toxin in the C-ST phage genome  
JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)  
MEDLINE 91024998

REFERENCE 2 (bases 36 to 2143)  
AUTHORS Kimura.K., Fujii.N., Tsuzuki.K., Murakami.T., Indoh.T., Yokosawa.N.  
and Oguma.K.  
TITLE Cloning of the structural gene for Clostridium botulinum type C1  
toxin and whole nucleotide sequence of its light chain component  
JOURNAL Appl. Environ. Microbiol. 57 (4), 1168-1172 (1991)  
MEDLINE 91282468

COMMENT These data kindly submitted in computer readable form by: Kouichi  
Kimura

Department of Microbiology  
Sapporo Medical College  
1 South, 17 West  
Sapporo 060  
Japan  
E-mail: nfujii@niguts.nig.ac.jp;pvax2.nlm.nih.gov  
Phone: 011-611-2111 x2334  
Fax: 011-612-5861.

## FEATURES

Source  
1..4592  
/organism="Bacteriophage c-st"  
/db\_xref="taxon:12336"  
255..260  
293..298  
305..310  
/note="SD sequence"  
321..4196  
/note="ORF for C1 toxin"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAAL4235.1"  
/db\_xref="GI:217781"  
/translation="MPITINNFNSDPVDNKNILYLDTHLNLANEPEKAFRITGNW  
VIPDRFSNSNLPKPPVTSKSGYDNPVSTDSKDTFLKELIKLFRKNSREI  
GELLYRSTDPDLPKNNVTPLNFDVDPNSVDVKTROGNNWVTKLSNPVILTG  
PRENIDETSTFKLINPFAQGGFALSIISPRFLIYSNATNDVGEFRSKE  
FCDPILLHLEHAMNLGIAIPNDQTLSSVTSNIFSYQNVKLYEIAEYAFGP  
ESSEGVYNNRKFVELYELQIFTEFNKAYINQVNRKIYLSNVYTPVNTILDDNV  
YIQGAFNIPKSNLNVLFMGONLNRNLPALRKVNENMLYFTKFKRAIDGRSLYKNT  
LDCRELLVNTDLPFGIDISDKTDIFLRKDINEETEVYIPDNVSDVQVILSKNTSE  
HQCQLLYPSIDSEELIPGENOVFDNRTONVYLSYXLYLESOKLSNDVDEPFTFR  
STEEALDNSAKYVTFPLANKYRAGVOGGLFLMWANDVDEFTNLRDITLKDTS  
VSAIPYGPALNISVRNGFTFAFVGTGVTILLFAPFTIPELFGALGVYISKVQE  
RNEITKIDNCLQKRWKQSYEMWGTWLSRIITQFNISYQMYDSLNYQAGATKA  
KIDLEYKYSGSDKENIKSOVENKSLDVKISEAMNNINKFIRECSVTYLFKNMLPK  
VIDELNEPRTNKAHLIDSHNLIIVGEVDKILKAKVNSFNQITFNIFSYTNKSL  
KIDLINEYNNNDKSLSLONKRNKTLVDTSYNAEYSEGDVQNLNIFPFDEKLGS  
GDRGKVIYQNTENIYVNSMYSESISFWIRINKVNSNLPGYTIIDSVKNSGWSIGI  
ISNLFYFLKQNEDESEQINSYDISNAPGNKWFVPTVTNNMGNNKVIYINGKLID  
TIKVELTGINSKFTTEINKPIDTGLITSDSDNNINMWIRDFYIFAKELDKDNLIL  
FNSLQYTNVYKDWGNDLRYNKEYYWNIDYLNRYMYANSRQIVFENTRRNNDFNEG

KIIIRKRGNTNTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYATCLR  
EQTRDINDNIIFIQIPNNNTYYAYKFSNFNGENISGICSTGTYRFLRGDWRHN  
YLPTVRQGNVSLLESTHGWFPVSE"

BASE COUNT 1869 a 460 c 714 g 1549 t  
ORIGIN

Query Match 48.5%; Score 665.2; DB 7; Length 4592;  
Best Local Similarity 68.3%; Pred. No. 1.5e-92;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTTCTCTACACCAACACTCCCTGTTGAAGGACATCAACAAC 72  
DB 2847 ACAATACCTTTAATATTTTTCATATACTAATATATCTTTATTAAAGATATATTAAT 2906  
QY 73 GAGTACTTCAACAACATCAACGACTCCAAGATCCCGTCCCTGCGAGAACCGGPAAGAACCC 132  
DB 2907 GAATATTTCAATAATTAATGATTCACAAAATTTTGACCTTACAAAACAGAAAAAATACT 2966  
QY 133 TTGGTCGACACTCCGGTTACAAACCGCGAGGTCTCCGAGGAGGTGAGTCCAGCTGAAC 192  
DB 2967 TTAGTGGATACATCAGGATATATATCGAAGTGAAGAGCGGATGTACACTTAAT 3026  
QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCGGTGAGGACAGAGGPAAGGTCAATC 252  
DB 3027 CCAATATTTCCATTTGACTTTAATTAGGTAGTTTCAGGGGAGGATAGAGTAAAGTTATA 3086  
QY 253 GTCACCCAGACGAGACATCGTCTACACTCCATGTACGAGTCTCTCCATCTCCATCTCCTC 312  
DB 3087 GTAACCCAGATGAATAATTTGTATATATTTCTATGATGAAGTTTGTAGCATAGTTTT 3146  
QY 313 TGGATCAGAAATCAACAAGTGGGTCTCCAACTTGGCAGGTTACACCATCATCGACTCCGTC 372  
DB 3147 TGGATTAGAATAAATAATGGGTAAAGTAAATTTACCTGGATATATATAATGTAGTGT 3206  
QY 373 AGACAACTCCGGTTGGTCCATCGGTATCATCTCCACTTCTCTGGTCTTCCACCTGAAG 432  
DB 3207 AAAATAACTCAGGTGGAGTATAGGTATTTAGTAATTTTTTTAGTATTTTACTTTAAA 3266  
QY 433 CAGACGAGGACTCCGAGCAGTCCCATCAACTTCTCTACGACATCTCCAAACGCTCCT 492  
DB 3267 CAAATGAAGATAGTGACAAAGTATAAATTTAGTTATGATATATCAATAATGCTCCT 3326  
QY 493 GGTACAAAGTGGTTCCTTCGTCACCGTCACCAACACATGATGGGTAAACATGAAGATC 552  
DB 3327 GGATACATAAATATGTTTTTCTTAACCTTTACTAACAATATGATGGGAAATATGAAGAT 3386  
QY 553 TACATCAACGGTAAAGTGTATCGACACCATCAAGGTCAAGGAGTTCACCGGTATCAACTTC 612  
DB 3387 TATATAATGGAAATTAATAGTACTATAAGTTAAAGAACTAACTGGAATTAATTT 3446  
QY 613 TCCAAGACCATCACCCTCGAGATCAACAAGATCCAGACACCGGTCTGATCACTCCGAC 672  
DB 3447 AGCAAACTATACATTTGAAATAAATAAATTCAGATACCGGTTTGTATTACTTCAGAT 3506  
QY 673 TCCGCAACATCAACATGTGGATCCGTGACTTCTACATCTTCGCCAAGGATTTGACCGGT 732  
DB 3507 TCTGATAACATCAATATGTGGATAGAGATTTTATATTTGCTTAAGAATTAGATGGT 3566  
QY 733 AAGGACATCAACATCCTGTTCACCTTCTTCAGTACACCAAGCTCGTCAAGACTACTGG 792  
DB 3567 AAAGATATTAAATATTTATTTAATAGTTGCAATATACTAATTTGTAAAAGATTTATGG 3626  
QY 793 GGTAAAGCTGTAGATACAAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGA 852  
DB 3627 GGAATGATTTAGATATATAAAGATATTTATATGTTTAAATATAGATTTTAAATAGA 3686  
QY 853 TACATACGCCAACTCCAGACAGATCGCTTCAACACAGACAGGTAAACAACAGACTTTC 912  
DB 3687 TATATGTATCGGAACCTCAGACAGAAATTTGTTTTAATACACGTAGAAAATAAATAAGTTC 3746  
QY 913 ACGAGGGTTACAGATCATCATAGCGGTATCAGAGGTATACACCAACGACACGAGATC 972  
DB 3747 AATGAAGGATATAAAATTAATAAAGAAAGATCAGAGGAAATACAAATGATACATAGAGTA 3806



```
QY 1033 ATGAAGACGAGACCATGTACGCGACACCACTCCACCGAGGACATCTAGCCATCGGT 1092
|||||
Db 4353 ATGAAGATGAACATATGATGACATATCATAGTACTGAAGATATATGCTATAGGT 4412
|||||
QY 1093 CTGCGTGACAGACCAAGACATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152
|||||
Db 4413 TTAAGAGAACAAACAAGATATAAATGATATATATATTTCAATACACCAATGAAT 4472
|||||
QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAAGGAGAGAACATCTCC 1212
|||||
Db 4473 AATACTTATATATACCATCTCAATATATTAATCAATATTTATATGGAGAAATATTTCT 4532
|||||
QY 1213 GGPATCTGTTCATCGATACCTACGATTCGCTCGGTGGTGACTGTCGACAGACAAAC 1272
|||||
Db 4533 GGAATATGTTCAATAGTACTTATCGTTTAGACTTGGAGGTGATGTTGATAGACAAAT 4592
|||||
QY 1273 TACTTGGTTCACACTCTCAAGCAGGTACTACGCTCCTCTCGTGAGTCCACTCCACC 1332
|||||
Db 4593 TATTGGTGCTACTGTGAAGCAAGAAATATGCTTCATTATTAGAATCAACATCACT 4652
|||||
QY 1333 CACTGGGAGTTCGTCACGACTCCCGAGTAA 1362
|||||
Db 4653 CATTGGGTTTTGTACTCTGAAGTGAATAA 4682
|||||

RESULT 4
CBCTOX
LOCUS
DEFINITION Botulinum bacteriophage genes for HA-17, HA-33, nontoxic components
and C1 toxin.
VERSION X62389
X62389.1 GI:558175
KEYWORDS botulinum toxin; C1 toxin; HA-17 gene; HA-33 gene; Hemagglutinin;
neurotoxin; nontoxic; toxin.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum;
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 9613)
AUTHORS Kimura,K., Fujii,N., Tsuzuki,K., Murakami,T., Indoh,T.,
Yokosawa,N., Takeshi,K., Syuto,B. and Oguma,K.
TITLE The complete nucleotide sequence of the gene coding for botulinum
type C1 toxin in the C-ST phage genome
JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)
MEDLINE 91024998
PUBMED 222445
REFERENCE 2 (bases 1 to 9613)
AUTHORS Tsuzuki,K., Kimura,K., Fujii,N., Yokosawa,N. and Oguma,K.
TITLE Nucleotide sequence of the gene for one of the components of
hemagglutinin produced by Clostridium botulinum type C
unpublished
JOURNAL 3 (bases 1 to 9613)
AUTHORS Tsuzuki,K.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept
of Microbiology, South 1 West 17, Sapporo 060, JAPAN
COMMENT On Oct 15, 1994 this sequence version replaced gi:40386.
See also X53041 & D90210.
FEATURES
Location/Qualifiers
1. 9613
/organism="Clostridium botulinum"
/strain="type C Stockholm, c-st"
/db_xref="taxon:1491"
complement(111..551)
/gene="HA-17"
complement(111..551)
/gene="HA-17"
/codon_start=1
/trausl_table=11
/protein_id="CAA44260.1"
/db_xref="GI:40387"
/db_xref="SWISS-PROT:P46083"
```

```
/translation="MSSERTFLPNQYKIKSLFSDSLYLYTSSGALSFSNTSLDNQK
WKLVSISSSNGFRFSNVAEPNKYLAVNDYGFIVLSSSSNNSLWNP*KIANSYIICLT
SVNVTYAWTIYDNNNTDQPTLNPDPEDINNSQILKLEK"
complement(560..565)
/gene="HA-33"
complement(614..1474)
/gene="HA-33"
/codon_start=1
/trausl_table=11
/protein_id="CAA44261.1"
/db_xref="GI:40388"
/db_xref="SWISS-PROT:P46084"
/translation="MSQTNANDLRNNEVFFISPSNNTKVLDKISQSEVKLWKLKLSG
NQKRLIYDNTKQAYIKVMDNTSLILITWAPLSSVSVKYDTNGDQOYWTLLQNYLSR
NVIIRNYPNLVLIYDITLAVTQTSSTSSNQFFKFSNCIYBALNRNCKIQTLNS
DRFLSKLNSQILVLMQWFSRQKWIIEYNKTSKAYTLKQENNYLTIWQINSYIV
ETVOSTSLIQYWNINYLNDNDASKYIILNQDTRVLDVYNSQIANGTHIVDSYHGN
TNQWMIINLI"
complement(1482..1486)
RBS 1726..1731
gene 1737..1737
CDS 1737..1737
/gene="nontoxic"
/gene="nontoxic"
/codon_start=1
/trausl_table=11
/product="non-toxic components"
/protein_id="CAA44262.1"
/db_xref="GI:40389"
/db_xref="SWISS-PROT:P46081"
/translation="MDINDLINSVPDKNVIVRAKINTFFKAFKVPAPNIWVAPE
RYGEPDIAEYKLDGIVDSNFSODSERENFLOAIILLKRINNTSGKOLLILI
STAPPVGYGGYSSPNFTGKTPKSNKLSLNTSTIPRPGGYRETNYESON
KNTFASNIIFGGSNIVENNVIYKKNDAENGMAASIVFQPLTIYKNFYDIP
AMETLKLIKSLYGLYKPSDNLVYRLTDLNKFQSLMIDLLISGGVDLEFI
NNTPYWTSNYSFNSIKMFKNYKTEYEGNAIGNDKLRKQKQINQVDIWNL
NLNPFQSFNSIIPRFSNALKHYRQYTYMDYTNNGINFGVNGQINTKPLSKN
TNIISKPKVVLNENNISLMKSNYIGDGLKGTEDFYTYKPYNEEYFNDSO
NPLNNTSIEVDSPRIIDNPKYDNLVFTQITSMTEVTTTALISNYLQAI
TNNEFTLSDFSKVSQKSLVYFSLNMLVLETIKNDGPIDQKYLWLVKSVF
KNYFDINTQELISMGINVLWFGKALNLTNSFVEEYQDSGASLISKDNL
REPIEDIDISLGLSFKDNLKLEIYISKNIYFKKIYFSLQWMTYYSQYFE
LIMAKQSILAESLVQIOWNKFTDUSKIPDILKIRETKTFFIDLSNESQIS
MRYVDNLNKAISICVFEVDIYKFSYMEKININIKITREFIQRTINQVLEKSLI
NSYFTKIDFELDIQIKFNSQVQVMEKILSPYLLFLAFKSPNSIIEIDSGK
NLTQYTESIELVGVNGESLYLAKSPNETIKFSNKFTGLTNFTICFWRFTGND
DKRLIGNKVNCGEIVFEDNGLVFRIIDSGNOBSVLSNIINDWYVITSVDRL
KQDLIFINDKNVANSIDOLISYINILSLVKNKNSIYSELVLDNPLEITSEVIR
NFFSLDNSYIRDSKSLLEKNKYNQILNVFFETSLYENDANKSYLSLKNYDGINI
SSVFKLINIDESKYVOKWDECTICVDGTETKLDISPENNRQILVSKDNKAKTIV
NTDLFRPCITFSYNDKYFSLSLRGGDYNMNCNDNNKVPKGAHLWLES"
5330..5335
5342..9217
/gene="C1 toxin"
5342..9217
/gene="C1 toxin"
/codon_start=1
/trausl_table=11
/protein_id="CAA44263.1"
/db_xref="GI:40390"
/db_xref="SWISS-PROT:P18640"
/translation="MPITINNFYSDPVNKNILYDLTLNLANPEKAFRTGNIW
VIPDRFSNPNKPNRVSQKGYDPNLYSTDSDKDTFLKEILKIFKRNSEI
GEELIYRSTDIPFGNNNTPTDFDVFNSVDVATROGNWNTKSGNSPVSITG
PRENIIDSTFTKLTNNFAAQGFALSIISPRFMTYNTATNDVGEGRFSKE
FCMDPILIMHLENAHNLGYIAIPNDQITSSVTNSIFYSQYNKLEYAIVAFGGP
TIDLIPKSARKYFEKALDYRSKRLNTTANPSFNKYGKOKILIRYFVY
ESSGEVTVNRKVELYNELTQJFETFNKIVNQRKLYLSNVTPTANTLDNRY
YDIQGNFNPKNLNVLFMGONLSEPNALPKVPEPNMLYLTKECHKAIDGRSLYKNT
LDCRELLVKNLTDLPFIGIDSVKTDIFLRKDINEETEVITYPNVSDQVILSKNSE
HGQLDLYPSDESEILPGENQVYDNRQNDVLYNSYILESKLSDNVEDFTTR
STEEALDSAKVYTYFTFLANKNAGVQGGFLFMWANDVVEDEFTILNRDLTKDIS
```

VSILPIVIGPALNISVRNGNEFAVGTGVTIIIELEAFPEFTIPALGAFVIYSKVOE  
 RNELITIDNCLPORKKNDSDVENMGWLSRIITOPNNISYOMVDSINYQAGAIKA  
 KIDLEFKYSSGSDKENIKSOVENLNSLDVKISBAMNINKEIRCSVTYIEKMLPK  
 VIDELNEEDNTAKLINLIDSHNILLVEGDVKAKVNSNPONTIPNIPFISYNNLS  
 LKDIINEFNINDKILSLQKRNLTVDTSYNAEVEBQVQLNIPFPDFKLIGSS  
 GEDRGKIVYQNEINVSMSYFSISFWIRINKVSNLPVYTIIDSVKNSGWSIGI  
 ISNPLVETLQNEQSDQSINFSVDISNAPGKNKFFVTVMNMKNKIYINGKLID  
 TIKVKELTGINFSEKTTTFEINKIPDGLITSDNINWIRDFYIPAKELDKDINIL  
 FNSLOVTVVVDYNGNDLYANKEYMYNDILYLRVYANSROIVENTRNNDFNEGY  
 KLIILKRLGNINTRVRGGDILYFDMTINKKAYNLNFMKNEIMYADNHSTEDIYAIGLR  
 EYKADINDIIFQIPQNMNNIYFASQIFSNFNGENISGICSGIRYRFLRGDMIRHN  
 YLPTVKQNGNYSLESTSTHWGFVPS\*

BASE COUNT 3846 a 1003 c 1348 g 3416 t  
 ORIGIN

Query Match 48.58; Score 665.2; DB 1; Length 9613;  
 Best Local Similarity 68.38; Pred. No. 1.4e-92;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTTCCTACACCAACATCCCTGTTGAGGACATCATCAAC 72  
 DB 7868 ACAATACCCCTTAATATTTTTCATATACATAATAATCTTTATTAAGAGATATAATTAAT 7927

QY 73 GAGTACTTCAACACATCAACGATCCCAAGATCTGCTCCGTCGAGACCGTAAGACACC 132  
 DB 7928 GAATATTTCAATATTAATGATTCAAAATTTTGACCTCAACACAGAAAAATATCT 7987

QY 133 TTGGTCGACACCTCGGTGTACACGCCGAGGTCTCCGAGGAGGTGACGTCACGCTGAAC 192  
 DB 7988 TTAGTGGATACATCAGGATATAATGACAGAGTGTGAGTGAAGAGCGGATGTCAGCTTAAT 8047

QY 193 CCAATCTTCCCATTCGACTCAAGCTGGGTCTCCCGTGGAGACAGAGGTAAAGGTCATC 252  
 DB 8048 CCAATCTTCCCATTCGACTTAAATAGTAGTGTACGGGAGGATAGAGTAAAGTTATA 8107

QY 253 GTACCCAGAACGAGAACATCGTGTACAACTCCATGACGAGTCCCTTCCTCCTCCTTC 312  
 DB 8108 GTAACCCAGATGAATATGATATATAATCTATGATGAAAGCTTTTAGCATTAGTTT 8167

QY 313 TGATCAGATCAACAGTGGTCTCCAACTGCGAGGTACACCATCATCGCTCGGTC 372  
 DB 8168 TGATAGATGAATAAATAGGTAGTAAATTTACCTGGATATACATAATATGATAGTT 8227

QY 373 AAGAACAACCTCGGTGTGTCATCATCTCCAACTTCCTGCTTCACCCCTGAAG 432  
 DB 8228 AAAATAAATCAGGTGGAGTATAGTATATTAGTAAATTTTATGATTTTACTTTAAAA 8287

QY 433 CAGACGAGGACTCCGACGAGTCCATCACTTCCTCAGACATCTCCACACAGCTCCT 492  
 DB 8288 CAAATGAAGATAGTGAACAAAGTATAAATTTTATGATGATATATCAAAATAATGCTCT 8347

QY 493 GGTACACAAGTGTCTTCCTGTCACCTGACCAACACATGATGGGTAAACATGAAGATC 552  
 DB 8348 GGATACAATAAATGCTTTTGTACTTCTACTAACAATATGATGGGAATATGAAGATT 8407

QY 553 TACATCAACGGTAACTGATGACACCATCAAGTCAAGGAGTTCACCGGTATCACTTC 612  
 DB 8408 TATATAATGGAATAATTAATAGATCTATAAAGTTAAAGTAACATGGAATAATTTT 8467

QY 613 TCCAACACCATCACCTTCGAGATCAACAGATCCGACACCGCTGTGATCACCTCCGAC 672  
 DB 8468 AGCAAACTATAACATTTGAATAATAAATTTCCAGATACCGGTTTGTACTTCAGAT 8527

QY 673 TCGGACAACTCAACATGTTGGATCGGTGACTTCATCATCTTCGCGCAAGGAGTTGGACGGT 732  
 DB 8528 TCTGATTAACATCAATATGTTGGAAGAGATTTTATATATTTTGTGTAAGAAATAGATGGT 8587

QY 733 AAGGACATCAACATCTGTTCAACTCTTCGAGTACACCAACGCTGGTCAAGGACTACTGG 792  
 DB 8588 AAAGATATTAATATATTTTAAATAGCTTGCATATATATATGTTGTAAGAAATATGG 8647

QY 793 GGTACGACCTGAGATACAAACAGGAGTACATGTTGTCACATCTGACTACTTGAACAGA 852

DB 8648 GGAATGATTTAAGATATAATAAAGAATAATTATATGTTAATATAGATTATTATAATAGA 8707

QY 853 TACATGTACGCCACTCCACAGATCGTCTTCAACACAGAGGTAAACACACGACTTC 912

DB 8708 TATATGTATGCGAACTCACCAGCAAAATGTTTTTAATACAGCTAGAAATATAATGACTTC 8767

QY 913 AACGAGGTTTACAGATCATCATCAAGCGTATCAGAGGTAAACACCAACGACACGAGTTC 972

DB 8768 AATGAAGATATAAATAATAATAAAGAATAACAGAGGAAATACAAATGATATAGTAGTA 8827

QY 973 AGAGTGTGACATCTCTGACTTCGACATGACTATCAACACAGGCGCTACACCTGTTTC 1032

DB 8828 CGAGGAGGAGATATTTATATTTTATGATGACAATAATAAACAAGACATATAATTTGTTT 8887

QY 1033 ATGAAGAACGAGACCATGTACGGCGACACCACTCCACGAGGACATCTACGCCATCGGT 1092

DB 8888 ATGAAGATGAACATATGTATGACAGATAATCATAGTACTGAAGATATATATGCTATAGGT 8947

QY 1093 CTGGTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCAATGAAC 1152

DB 8948 TTAAGAACAAACAAGAGATATAAATGATATAATATATTTTCAATACACCAATGAAT 9007

QY 1153 AACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC 1212

DB 9008 AATCTTATTTATGACATCTCAAAATTTTAAATCAAAATTTTAAATGAGAAAAATTTCT 9067

QY 1213 GGTATCTGTTCCATCGGTACTACAGATTCCTGCTGGTGGTGACTGGTACAGACACAA 1272

DB 9068 GGAATATGTTTCAATAGTACTTATGCTTTAGACTTTGGAGGTGATTGGTATAGACAAAT 9127

QY 1273 TACTTGGTTTCAACTGTCAAGCGGTAACTACGCTCTCTGCTGGAGTCCACTTCCACTC 1332

DB 9126 TATTTGGTGCCTACTGTGAAGCAGGAATATATGCTTCATTATTAGTAATCAACATCAACT 9187

QY 1333 CACTGGGATTCGTCACGATCTCCGAGTAA 1362

DB 9188 CATTTGGGTTTTGTACCTGAAGTGAATAA 9217

RESULT 5  
 CBP1CNTC1 Clostridium botulinum phage 1C, Chn-14, Chn-33, Chn-138 and BONT/C1  
 LOCUS 9689 bp DNA linear BCT 15-FEB-1999  
 DEFINITION Clostridium botulinum phage 1C, Chn-14, Chn-33, Chn-138 and BONT/C1  
 genes.  
 ACCESSION X66433.1 GI:509271  
 VERSION X66433.1  
 KEYWORDS haemagglutinin; neurotoxin.  
 SOURCE Clostridium botulinum.  
 ORGANISM Clostridium botulinum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 REFERENCE 1 (bases 1 to 9689)  
 AUTHORS Hauser, D.F., Eklund, M.W. and Popoff, M.R.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 9689)  
 AUTHORS Hauser, D.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-1992) D.F. Hauser, Institut Pasteur, Unite des  
 Toxines Microbiennes, Institut Pasteur, 28 rue duDocteur-Roux,  
 F-75724 Paris Cedex, FRANCE  
 FEATURES  
 source  
 1. 9689  
 /organism="Clostridium botulinum"  
 /db\_xref="taxon:1491"  
 complement(253..647)  
 /gene="Chn-14"  
 complement(253..633)  
 /gene="Chn-14"  
 /codon\_start=1  
 /transl\_table=11  
 /product="14kDa protein associated with BONT  
 /C1-haemagglutinin complex"  
 /protein\_id="CAA47057.1"

```

/db_xref="GI:509272"
/db_xref="SPTREMBL:Q45915"
/translation="MSSETEFLNGNKKIKLSFSDSLYLYTSSGALSFQNTSSLDNKK
WKLEYISSSNGFESNAENKYLAYNDYGFVLSSSNSLNWPKIKALNSYIICTL
SIVNYDYATIWDNNNNIDHLE"
complement(642..647)
/gene="Chn-14"
complement(696..1556)
/gene="Chn-33"
complement(696..1556)
/gene="Chn-33"
/codon_start=1
/trans_table=11
/product="33kDa protein associated with BoNT"
/Cl-haemagglutinin complex"
/protein_id="CAA47058.1"
/db_xref="GI:509273"
/translation="MSTQTNANDLRNNEVFIPSSNNTNKVLDKISQSEVKLNKLSGA
NOKRLIYDNRKQAKYIKVMDNTSLITLNNAPLSSSVKTDINGNOYKYLONYSR
NVIIRTNMNNLVQYINIDTDLAVSTQTSNNOFFPSCNIEALNRNCKQOTQLNS
DRFLSKNLNQLIIVLWQFSSRSKQWIEETNFKSATYLLKQENNRILYTIQNSNNY
ETYSTSLQIYWNLYLDNDASKYILYINLQDTRVLDVNSQIANGTHIVYDVSYGH
TNQWIIINLI"
complement(1564..1568)
/gene="Chn-33"
complement(1564..1568)
/gene="Chn-33"
join(1803..1808,1819..5409)
/gene="Chn-138"
1803..1808
/gene="Chn-138"
1819..5409
/gene="Chn-138"
/codon_start=1
/trans_table=11
/product="138kDa protein associated with BoNT"
/Cl-haemagglutinin complex"
/protein_id="CAA47059.1"
/db_xref="GI:4375225"
/translation="MDINDLNINSPVDNKNVIVRARKNTNTEFFKAFVAPNIWAPE
RYGEPLDIAEYKLDGIVDSNPLSDSRENFLOAIILIKRINTTISGQLLSLI
STALFFYIGGGISFNPFTFKPKSNKKNSLVTSTIPFPFGYRETNIESQN
KNKYSVILFGGSNIVNNVYIKKDAENGMTMAEIVFPQLTYKRYKFIIDP
AMELTKLISLYGKIPSDNLWPEYRLTELDNKPQSLNIDILLISGGVDEFI
NTNYPWTNSYFNSIKMFQKNIYKTEGNAIGNDIKRLKOKFOINQVDIWNL
NLNFKOSFNSIIIDRESNALKHRYKQYTYMDYTDNYINGFVNGCINIKPLSNK
TNILSKPKVNLVNNENISLMKSNYIGDLKGTEDFYSTYKIPYNEEYEFNDSD
NPLNLSIEPDSIPELIIDNPYKONSNLVETQIISMTTEYTHALSINYLQAI
TNNTFLSSDFSXVWSKOKSLYISLQNLNLSLETIKNDGFDIDDKIYILWKEVF
KNYSPDINLQEDISMCGINEVLFQKALNINLTNSFVEEYQDSGALSLSKDDNL
REPNEIDDISLGLSFKDLNNKLYEISKNIVFKRIYFSLDQWTEYISOYE
LICMAKQSLAQESLVQIKVQNFDTLSKASIPDPLKIRETEKTFIDLNGESQIS
MNRVDNFKAKSICVFVEDIYKPIFSYMEKYINNINIKREFTORCNTINDNEKSILI
NNTYFETIDRFLOISIKNFNSQVQVMEKILSPVOLLFASKGPNINILEDISGK
NTLIQYFESIELVYGNGESLYLKSPNETIKESNKEFTGLTNNTFCWFLPFGKND
DKTRLIGNKNCWGEIYFEDNGLVFEIISNGNSQVLSNLSINDNNWYISVDRL
KQDLILINDKNVANVSDQLLSYISNIIISLVNKNNSIYVEELSVDNPIITSEVIR
NYFSLDYSYIRDSKSLLENKYNQLYNFFPETSLEYEVNKNKSLYLSKNTDGINI
SSVFKLINIDESYVYQKWDCEICIVLDGTEKYLDISPPENNRILQVLSKDNAKTIV
NTDLFRPDCITFSYNDKYFSLSLRDGYNMNCNDNKNKPKGAHLILES"
5408..5413
/gene="BoNT/Cl"
5408..5413
/gene="BoNT/Cl"
5424..9299
/gene="BoNT/Cl"
5424..9299
/gene="BoNT/Cl"
/codon_start=1
/trans_table=11
/product="Botulinical Cl neurotoxin"
/protein_id="CAA47060.1"
/db_xref="GI:509275"

```

```

/db_xref="SWISS-PROT:P18640"
/translation="MPTIINNFNSDPVDNKNLILDTHTLIANEPEKAFRITGNIW
VIPDRSRNPNLNKPPRVTSKGYDPNYSLSDSDKDTLKEIILKLFKRSREI
GEELIVRLSDIIPGNNNTPIPTFDVDNFSVDVKTRGNNWVKTSINPFSVIITG
PRENIDPETSFKLNNTPAEGEGSALISISPREMLTYSNATNDGEGFRFSKSE
FCMDPILILMHELNHAMNLYGTAINDOTISSTNIIYSQYNKLEVAEYIARGGP
TIDLPKSRKYFEERKALDYIRAKRLNSITTPANFSSNKYIGEKQLIKRIRFVV
ESSGEVTNRNKFELYNELTQITFEFNKAYINQNRKYLSNYPVPTANILDDNV
YIQNGENIPKSNLNVLMGONLRNPAFKVNPENMLYFTKCHKAIDGRSLNKT
LDCELLVKNTDLPFGDISDKTDIFLRKDINEETEVIYVNDVSDQVILSKNTSE
HGOLDLPSIDSESEILPGENOVFYDNRQNVYLSNYSYLESQKLSNDVEDFTFR
SIEBALNSAKVYFPTLANKYAGVQGLFLMWANDVVEDETTNLRKDLTKLSD
VSAIPIGPAIANSYVRGNETEAFAYVTGVTILLPEPEPTIPALGAFVIYSKVOE
RNEITIKTIDNLSQRKRWKSDYEMWGWLSRIITQFNNSYOMKDSLNLYQAALKA
KIDLEYKYSKDRKIKSQVENLSDVKISEAMNNLNKPIRESCTVYLFKNMLPK
VIDELNEFDRTAKLINLDSHNIILVDEVDKAKVANSFQNTIPFNIFYTNNSL
LKDINEFYNNINDSKILSLQNRNLTVDTSYNAEVSSEGDVQLNPIFFPFLGSS
GDEGRKVIYNNENIVNYSFISFWRIRNKWYSNLPGVTIIDSVKNNSSQWSTGI
ISNPLVTLAKQNSDSQISNFSYISNAPYKNWFEVTVNMNMKMYIYNGKLID
TIKVKELGTNFSKTTTFEINKIPDGLITSDSDNIMKATROEYIFAKELDGKDNIL
FNSQIYTNVYDIDWGNDRINKKEYYVWNIIDYLNRYMYANSRQIVFNRNNDNFEGY
KILIRGNTNDIRVGGDILYFDMTNNKATNLFMKNETMTADNHSHTDIATGLR
EQTDINDNTIFQIOPMNTYYASQIFKSNFNGENISGICISGTYRFLGGDWYRHN
YLVPTVQGNVYASLLESTSTHWGFVPVSE"
9385..9424
/gene="BoNT/Cl"
9385..9424
/gene="BoNT/Cl"
/Note="putative"
BASE COUNT 3865 a 1022 c 1359 g 3443 t
ORIGIN
Query Match 48.5%; Score 665.2; DB 1; Length 9689;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
QY 13 ACCATCCCATCAACATCTTCTCTACACCAACACTCCCTGTTGAAGGACATCATCAAC 72
DB 7950 ACAATACCCCTTAATATTTTTCATATACTATAATCTTTATTATAAGATATAATTAAT 8009
QY 73 GAGTACTTCACAAACATCAACGACTCCAGATCTCTCCCTGCAGAACCGTAAAGAACCC 132
DB 8010 GAATATTCATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8069
QY 133 TTGGTCGACACCTCCGGTACAAACCCGAGGTCCTCCGAGGAGGTCGAGCTCCAGCTGAAC 192
DB 8070 TTAGTGGATACATCAGGATATAATCAAGTGAAGTGAAGGCGGATGTCAGCTTAAT 8129
QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTTCTCCGGTGAGGACAGAGGTAAAGTCTATC 252
DB 8130 CCAATATTTCCATTTGACTTTAAATTAGGTAGTTCAGGGGAGGATAGAGGTAAAGTTATA 8189
QY 253 GTCACCCAGAACGAGAACATCGTCTACAACTCCATGTACGAGTCTCTCTCATCTCTCTTC 312
DB 8190 GTAACCCAGAAATAATATGATATATAATTAATTAATTAATTAATTAATTAATTAATTAAT 8249
QY 313 TGGATCAGAAATCAACAGTGGGTCTCCAACTTCCAGGTGTACACCATCATCGACTCCGCT 372
DB 8250 TGGATTAGAAATAATAATAGGTAAAGTAAATTTACCTGGATATACATAATGATAGTGT 8309
QY 373 AAGAACACTCCGGTGTGTCCTCATCGGTATCATCTCCAACTTCTCTCTCTCCCTCAAG 432
DB 8310 AAAATACTCAGGTGGAGTATAGGTATATTAGTAATTTTTTAGTATTATTACTTTTAAA 8369
QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCCACACAGCTCCT 492
DB 8370 CAAAATGAAGATAGTGAACAAAGTATAAAATTTAGTTATGATATATCAATAATAGTCTCT 8429
QY 493 GGTTCACAACTGCTTCTCGTCACCGCTCCACACACATCATGGGTGATACATCAAGATC 552
DB 8430 GATACATAATAATAGTGTGTGTGTAATCTTACTACATAATGTTGGGAATATAGAGATT 8489
QY 553 TACATCAACGGTAAAGTGTATGATCGACACCATCAAGGTCAAGGAGTTCACCGGTATCACTTC 612

```

Db 8490 TATATAAATGAAAAATTAATAGACTACTATAAAAGTTAAAGAACTAACTGGAATTAATTT 8549  
QY 613 TCCAGACCATCACTCGAGATCAAGAAGATCCAGACACCGGTCTGATCACTCCGAC 672  
Db 8550 ACCAAACTATAACATTTGAATTAATAAATCCAGATACCGGTGATCTTCAGAT 8609  
QY 673 TCCGACACATCAACATTTGGATCGGAGCTTCACATCTCCGACGAGGTGGACGTT 732  
Db 8610 TCTGATACATCAATATCTGGATAAAGAGATTTTATATATTTGCTAAAGAAATPAGATGT 8669  
QY 733 RAGGACATCAACCTCTCTCAACTCTTCAGTACACCAACGTCGTCAGGACTACTGG 792  
Db 8670 AAGATATTAATATATTAATTAATAGCTTGCATTAATCTATGTTGTAAGAATTAATGG 8729  
QY 793 GGTACGACCTGGAGATACACAAAGAGTACTACATGTGCAACATCGACTACTTCTGAACAGA 852  
Db 8730 GGAATGATTTAAGATATAATAAAGAAATTAATATGTTTAAATAGATTAATTAATAGA 8789  
QY 853 TACATGTAGCGCACTCCAGACAGATGCTCTCAACACCAAGCTACACCAACGACTTC 912  
Db 8790 TATATGTAGCGCACTCCAGCAATGTTTAAATACAGTAGAATAATATATGACTTC 8849  
QY 913 AACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGGTAAACACCAAGCAGACGAGTC 972  
Db 8850 AATGAAGATATAAAATTAATAAAGAAATCAGAGAAATACAAATGATAGTAGTA 8909  
QY 973 AGAGTGGTGACATCCTGCTACTTCGACATGACTATCAACACAGGCTACACCAACGCTTC 1032  
Db 8910 CGAGGAGGAGATATTTATTTGATGATCAATTAATCAAAAGCATATAATTTGTTT 8969  
QY 1033 ATGAAGAAGGAGACCATGTAGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT 1092  
Db 8970 ATGAAGANTGAACATATGTATGCAGATTAATCATAGTACTGAAGATATATATGCTATAGT 9029  
QY 1093 CTGCGTGACACCAACCAAGGACATCAACACCAACATCATCTCCAGATCCAGCCAAATGAAC 1152  
Db 9030 TTAAGAGAACCAACCAAGGATATAATGATATAATATATATTTCAAAATCAACCAATGAAT 9089  
QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212  
Db 9090 AATACATTAATTAACGATCTCAATATTTAAATCAAAATTTAATGAGAAATATTTCT 9149  
QY 1213 GGTATCTGTTCATCGGTACCTAGATTCGCTGGTGCTGGTGCTGCTAGACACCAAC 1272  
Db 9150 GGAATATGTTCAATAGGTACTTATCGTTTAACTTGGAGGTGATGTTGATAGACAAAT 9209  
QY 1273 TACTTGGTTCCTCACTGTCAGGAGGTAACTACGCTCTCTGCTGGAGTCCACTTCCACC 1332  
Db 9210 TATTGGTGCCTACTGTGAGCAAGAAATTAATGCTTCAATTAATAGATCAACATCAACT 9269  
QY 1333 CACTGGGGATTCGTCCTCCAGTCTCCGAGTAA 1362  
Db 9270 CATTGGGGTTTTGTACCTGTAAGTAAATA 9299

RESULT 6  
AB061780 11747 bp DNA linear BCT 06-NOV-2001  
LOCUS  
DEFINITION Clostridium botulinum orf-22, ha-70, ha-17, ha-33, ntaha, nt genes  
for ORF-22, HA-70, HA-17, HA-33, NTNHA, neurotoxin, complete cds.  
ACCESSION AB061780  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Clostridium botulinum (strain:C-Voichi) DNA.  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
Clostridium.  
REFERENCE 1  
AUTHORS Sagane,Y., Kouguichi,H., Watanabe,T., Sunagawa,H., Inoue,K.,  
Fujinaga,Y., Ogura,K. and Ohshima,T.  
TITLE Role of C-terminal region of HA-33 component of botulinum toxin in  
hemagglutination

Biochem. Biophys. Res. Commun. 288 (3), 650-657 (2001)  
21534265  
2 (bases 1 to 11747)  
Sagane,Y., Watanabe,T., Kouguichi,H., Morita,M., Miyamoto,T.,  
Yamamoto,T., Muroga,A. and Ohshima,T.  
Direct Submission  
Submitted (16-MAY-2001) Toshihiro Watanabe, Tokyo University of  
Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido  
099-2493, Japan (E-mail:t-watanabe@bioindustry.noda.ac.jp,  
Tel:81-152-48-3843, Fax:81-152-48-3843)  
Location/Qualifiers  
1. 11747  
/organism="Clostridium botulinum"  
/strain="C-Voichi"  
/db\_xref="taxon:1491"  
6. 545  
/gene="orf-22"  
/gene="orf-22"  
/codon\_start=1  
/transl\_table=11  
/product="ORF-22"  
/protein\_id="BAB71744.1"  
/db\_xref="GI:16380754"  
/translation="MNDLFYAIENLKHNDNQNFNIEMSLKKYIETSKKYNLYDYNN  
DILYHLMKEIENLNFSELDLRKYSTIKRYCINCKKNDKIIINSEVYIK  
KLDAVNVSLYCDNFELDLISILNYKEQIIYMKFEFGKDKNEAIRLRSQSIYK  
IRUKSLKLYPIVMQLVNI"  
/complement(620. 2491)  
/gene="ha-70"  
/complement(620. 2491)  
/gene="ha-70"  
/codon\_start=1  
/transl\_table=11  
/product="HA-70"  
/protein\_id="BAB71745.1"  
/db\_xref="GI:16380755"  
/translation="MSLSIKELIYTKDSINNVLADGNVYVNRGDGWLISRONNLG  
GNISNGCTAIVGLRIRETATPIYPTAFNEEIKKNVQVNFANFEASIPGFE  
FSKTPAKSLKLYQTYIRYEIKVLQNTVERAVLYVPSLVYKSEFENSEQID  
KNFTQSDKCIENKEFIYKIDDTIVKESNNNNINFTSOTILPYNGLYINK  
CDGYMNTDKDLIGTLIESSTGSIQPLRNTTPTLNTSPTIFSOEYTEARLND  
AFQLENTSTLTFKVEEATNKINISMYNTYKELINTQNGINDKAPYLP  
GKCEVSPSPQAPVETPVDDGFIQTGNENIIVGVINPSENIEESTPIPDQYTY  
NIPISQNNACVLFVNTGTYKIKTKNIPLIIEAIGSNRNNSNNISNDIK  
AIRITGKNSDKSYLIVSLFKDKNYIRIPIQSSSTSQLFKRELNSLDADST  
VNLDNLSTGTHYTRQSPDVNGYISYQITPDGFNNIASSIFSPTRNNQIGTLY  
RLTESINGYLIITNNYSDDLNVPEISLNGATYIFRVKVTLENNYNIIFDAIRNS"  
/complement(2505. 2945)  
/gene="ha-17"  
/complement(2505. 2945)  
/gene="ha-17"  
/codon\_start=1  
/transl\_table=11  
/product="HA-17"  
/protein\_id="BAB71746.1"  
/db\_xref="GI:16580756"  
/translation="MSSRTFLPNGYKIKSLFNSLIYLYSSGALSFSNTSLDNQK  
WKLEIVSSNGFSPENVAEPKYLAVNDYGFIVLSSSSNNLWNPKIANSYIICLT  
SIVNVDYATIIDNNNIIDQILNLPNFIDINNSQILKLEKL"  
/complement(2999. 3856)  
/gene="ha-33"  
/complement(2999. 3856)  
/gene="ha-33"  
/codon\_start=1  
/transl\_table=11  
/product="HA-33"  
/protein\_id="BAB71747.1"  
/db\_xref="GI:16580757"  
/translation="MSQTNANDLRNNEVFTSPSNNNTKVLKDISSEVKLNNKLSGA  
NQKRLIYDTNKQAYKIKVMDNLSILTNAPLSSSVKTDNGDQYWLQNYISR  
NVILRNPNPLQYKNDTLMVSTOTSSNOFFKFSNCIYESFNNSCKIGTSLTI  
KFIKQNSNNVTWSNNGDNQKWKLYNESKMAVTLTCTIKKNKNEIWFESSIGNVVG

```

TYRTGNDQYWFNYLNNDASMYTISNFSNQSKFLDWNGLADGTNVQWVDSNGTS
aokwiitrl"
4119..7709
/gene="ntnha"
4119..7709
/gene="ntnha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAB71748.1"
/db_xref="GI:16580758"
/translation="MDINDLNINSPVDNKNVIVRARKNTNTFFKAFKVPAPNIWPAE
RYGPLEDAIEYGLSDIYDLSQDSRENFLQAIILLKRINNTISGKQLSLI
STAIPFYGYGGYSGNIYFTGKTSKSNKLSLVTSTIPFPFGYRENTYIESQ
KNMFYASINIVFGSGNTVENNYKKNDNAENGMTMAEIVFQPLITYKNKYFIDP
AMELKILKSLFLYG:KPSDNLVWPRLTELDNKOFSQNLIDLLISGGYDLFI
NPNYFNTSVPNSIKMFEXYKNLYKTEIEGNAIGNDKLRKQFQINVOIDNWL
NLNYFCSFNSIIPDRESNALKHEKRYQYTMQYTDYNNINGPVGQINIKLPSKN
TNIIKPKYVNLVNNNISLNKSNITGDGLKGTIEDFYSTIKIPNEYEEFNSQD
NPNLNIIEVDISPEIIDINPYKDSNDLVFTQITSMTEEVVTHTALSINLQAI
TNNEFTLSDFSKVYSKDSLYSFLDNLMSYLETIKNDGPIIDTKYILWLKEV
KNYSPDINTQIDSMGNEVWLFKALNINLNTSINSFVEEYQDSGALISLKKDNL
REPNIIDDISLGLSGFKDNLNKLVEIKSNIVYFKIYFSLDQWTEYYSQFPE
LTCMAKQSLAQESVKQIVONKFTDLSKASIPDDTLKIRETEKTEFIDLSNESOIS
MKNVDFNLKASICVFEDYIPKISYWEKVIYNNINIKTREFTORCTININDKESLI
NSYFTKIDFALDQSKNFSQVQVMEILSPIQLLFLASKGPNLSNIIEDSGK
NTLIQYTESIELVGVNGESDYLKSPNETIKFSNKFTNGLTNNFTICFWLRETND
DTRLIGHKNVNGWEYFEDNGLVFIIDSGNQESVYLSNINDNMYIYSISVDR
KDQLIFINDKNVNSIDQLSIYSTNIIISLVNKNNSIYVEELSVLDNPITSEEVIR
NYSYLDNNSIRDSKSLKYNKLYNVYFPEYSLYEVNDNKNKSVLSLKNQDINI
SSVKFLKINIDESKVVQWDECIICVLDTGTEKYLDISPENNRITQVSSKNKKTIV
NTDLFRPQIIFSYNDKYFSLSLRDGYNMHCNDNKNKPKGAHLWLES"
7724..11599
/gene="nt"
7724..11599
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BAB71749.1"
/db_xref="GI:16580759"
/translation="MPITINNFNSDPVDNKNILYDLTHLNTLANEPEKAFRITGNW
VTPDFNSNPNKPPVTSXPXGYDPNLSDSDKDTFLKEILKFKRINSREI
GELIYRLSDIPDPGNNTPIITDFDVFNSVDVKTROQNNWVKTGSINPSVIIT
FENDIDPETSFKLTNTFAAQFGFALSII:SPRFMLYSNATNDVGEYFQSKSE
FCMDPILILHMLHANMELXGAIIPNDOTISSVTSNIFYSQYNVLEYAIEYAFGS
TIDLPKARKVEEKALDYRSIAKRLNSITANPSFENKYGEGYKQKILKRYRW
ESSGEVYVRNKKFVELINELIQITERNAKIYNQKRYLSNVITPVTANLDDNV
YDIQNGFIPKSNLVLPFGQNLSPALRKVPENMLYLFTEKCHKAIDRSLYNT
LDCRELLVNTDLDPFIDIGSDVKTDFLRKDINTEEVYYPDNVSVDVILSKNTE
HQLDLPLPSIDSEIILPGNQVFNQNTQNDVLSYVYLESQKLSNDVEDFTIR
STEEALDNSAKYIVFPTLANKVAGVGGFLMWANDVVEDFTNLRKDTLDKISD
VSAIIPYIGPALNINSVRGNFTEAPVGTVILLEAFPEFTIPALGAFVIYSKQE
RNEIITDNCLEQRIKWKDSYEWMTWLSRIITOPNKSIXQMYDSLVYQAGAIKA
KIDLEKYSQSDRENISQVENIKNSLDYKISAMANNKIFIRECVSYFLKNNLPK
VIDELNEPDRNKAHLINDSHNIIIVGEVDKUKARKVNSFQNTIPFNISFNLSL
LKDIINFEFNINDSKLSLQNRKNTLDVTSQYNAEVSDEGDVQLNIPFPDKLSS
GDRGKVIYQNTQENYSMSYESFISFWIRKNVSNLPGYTIIDSVKNSWSGICI
ISNVLFLTKQNEDESEQINFSDIISNAPGNKFWFTVNNMNMKVIYINGKILDI
TIKVELTGKFSKTIITPEINKIPDGLTSDSDINNMTRDFYIPAKELDKDINIL
FNSLOYVNVQWYDNGNDLRYNKYMYNIDLYNMYANSROYFVNTRRNNDNFGY
KLIIRIRGNTNDRVGRGDILYEDMTINNKAYNLPKMKNETMYADNHSTEDYALGR
EOTKNDINILFIQIPMNTIYISQIFKSNFNGENISGICSIQTYFRUGGDWYRHN
YLPVPKQGNVYASLLESTSTHWGPVSE"
BASE COUNT 4574 a 1263 g 1622 g 4288 t
ORIGIN

```

```

Query Match 48.5%; Score 665.2; DB 1; Length 11747;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
13 ACCATCCCATCAACATCTCTCTACACCAACAACACTCCCTGTTGAAGGACATCATCAAC 72
|||||

```

```

Db 10250 ACAATACCCCTTTAATATATTTTTCATATATCTTATTAATCTTTTATTAAGATATATAAT 10309
QY 73 GAGTACTTCAACAACATCAACGACTCCAAGACTCTGTCCTCTGAGAACCGTAAGAACACC 132
Db 10310 GAATATTTCAATAATATTAATGATTTCAAAAATTTTGACCTTACAAAACAGAAAAATACT 10369
QY 133 TTGGTCGACACCTCCGGTTTACAAACCCGAGGTCCTCCGAGGAGGTGAGCTCCAGCTGAAC 192
Db 10370 TTAGTGGATACATCAGGATATTAATGCAGAAAGTGAAGTGAAGGCGGATGTTCAAGTTAAT 10429
QY 193 CCNATCTTCCCATTCGACTTCAAGCTGGGTCTCCCTCGGTGAGGACAGAGTAAGCTGATC 252
Db 10430 CCAATATTTCCATTTGACCTTTAAATAGGTAGTTCCAGGGGAGGATAGGTAAGTATA 10489
QY 253 GTCACCCAGAACGAGAACATCGCTCAACACTCCATGTACGAGTCCCTTCCTCATCTCCCTTC 312
Db 10490 GTAACCCCAATGAAATATTTGATATATTAATCTATGTAAGAAAGTTTGTAGCATTAGTTT 10549
QY 313 TGGATCAGAAATCAACAAGTGGGTCTCCAACTTCCAGCTTACCACTCATCGATCCGCTC 372
Db 10550 TGGATTAGAAATAAATAATGGTAAAGTAAATTTTACCTGATATATACTATAATAGTAGTGT 10609
QY 373 AAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACCTTCCCTGCTTCCACCTCAAG 432
Db 10610 AAAAAATCTCAGTTGGAGTAGTAGTATTTATAGTAATTTTGTAGTATTTTACTTTAAA 10669
QY 433 CAGAAGGAGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTTCCAAACGCTCCCT 492
Db 10670 CAAAATGAAGTAGTAGAACAAAAGTAAATTTTAGTTATGATATATACTATAAATAATGCTCCT 10729
QY 493 GGTTCACAAAGTGGTCTTCTGTCACCGTCCACCAACAATGATGGGGTAACATGAAGATC 552
Db 10730 GGATCAATTAATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10789
QY 553 TACATCAACGGTAAAGTGTGATCGACACCATCAAGTCAAGGAGTTGACGGGTATCAACTTC 612
Db 10790 TATATAATGGAATAATTAATAGATATATAAAGATTAAGAATACTGGAATAATTTT 10849
QY 613 TCAAGACCATCACCTTCGAGATCAACAGATCCAGACACCGGTCGATCACTCCGAC 672
Db 10850 AGCAAACTATTAACATTTGAAATAAATAAATCCAGATACCGGTTGATTTACTTCAGAT 10909
QY 673 TCCGACACATCAACATGTGGATCCGTGACTTCTACATCTTCTGCAAGGAGTTGAGCGGT 732
Db 10910 TCTGATAACATCAATATGTGGATTAAGAGATTTTATATATTTGCTAAAAGATTAAGTGGT 10969
QY 733 AAGGACATCAACATCCTGTTCACCTTTCAGTACACCAACGTCGTCAGGAGTACTGG 792
Db 10970 AAAGATTAATATATATTTAATGCTTTGCAATATATACTATGTTGTAAGAGATTTATGG 11029
QY 793 GGTAAAGACCTGAGATACAAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGA 852
Db 11030 GGAATGATTTAAGATATAATAAAGAAATATATATGTTAATATAGATTTATTAATAGA 11089
QY 853 TACATGACCCACTCCAGACAGATCGCTTTCACACACGACAGCTAAACAACAGACTTC 912
Db 11090 TATATGATGGAACCTCAGCAAAATGTTGTTTAATACCGTAGAATAATAAATGACCTTC 11149
QY 913 AACGAGGTTTACAAGATCATCATCAAGCGTATCAGAGGTAAACCAACGACACAGAGTC 972
Db 11150 AATGAAGGATATAAATAATAAATAAAGAAATCAAGAGAAATACAATGATAGAGTA 11209
QY 973 AGAGTGGTGACATCTCTGACTTCAGATCATCAACAACAGGCTTACACCTGTC 1032
Db 11210 CGAGGAGGAGATTTTATATTTTGTATGCAATTAATAAACAAGCATATAATTTGTTT 11269
QY 1033 ATGAAGAAGGAGACCATGTACCGCAGCAACCACTCCACGAGGAGCATCTAGCCATCGGT 1092
Db 11270 ATCAAAATGAACATATGATCGAGATAATCATAGTACTGAGATATATATGCTATAGT 11329
QY 1093 CTCGGTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCAATGAAC 1152
Db 11330 TTAAGAGAAACAAACAAAGGATTAATGATGATATATATTTTCAATATACACCAATGAAT 11389

```



```

QY 1153 AACACTTACTAGCTCCAGATCTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212
DB 11390 AATCTATTATTAGGATCTCAATATTATTAATCAAAATTTATGGAGAAATATTCT 11449
QY 1213 GGTATCTTCCATCGGPACCTACAGATTCGGTGGTGGTGGTACGACACAC 1272
DB 11450 GGAATATCTCAATAGGTACTATCGTTTATAGACTTGGAGGTGATTGGTATAGACACAAT 11509
QY 1273 TACTTGGTCCAACTGTCAACGAGGTAAGTACGCCCTTGTGAGTCCACTCCACC 1332
DB 11510 TATTGGTCCCTACGTGACGACGAGGAATATGCTTCATATTAGATCAACATCAAT 11569
QY 1333 CACTGGGAGTTCGCCAGTCTCCGAGTAA 1362
DB 11570 CATTGGGTTTGTACCTGTAAGTGAATAA 11599

RESULT 7
LOCUS CECBONT
DEFINITION Clostridium botulinum C phage BONT/C1, ANTP-139, ANTP-33, ANTP-17, ANTP-70 genes and ORF-22.
ACCESSION X72793
VERSION X72793.1 GI:516171
KEYWORDS Antp gene; bont gene; DNA binding protein; neurotoxin type C1.
SOURCE Clostridium botulinum phase 1C.
ORGANISM Clostridium botulinum phase 1C
          Viruses.
REFERENCE 1 (bases 1 to 12297)
          Hauser,D., Eklund,M.W., Boquet,P. and Popoff,M.R.
          Organization of the botulinum neurotoxin C1 gene and its associated
          non-toxic protein genes in Clostridium botulinum C 468
          Mol. Gen. Genet. 243 (5), 631-640 (1994)
          94301293
          MEDLINE
          PUBMED 8028579
REFERENCE 2 (bases 1 to 12296)
          Hauser,D.F.
          Direct Submission
          Submitted (03-MAY-1993) D.F. Hauser, Inst. Pasteur, Unite des
          Toxines Microbiennes, 28, Rue du Docteur Roux F-75724, Paris Cedex
          15, FRANCE
          On Jul 27, 1994 this sequence version replaced gi:509268.
          Related sequences: X53751 & X66433.
FEATURES             source
     source
     1..12297
         /organism="Clostridium botulinum phase 1C"
         /strain="468"
         /sub_species="type C"
         /db_xref="taxon:35348"
         /lab_host="Clostridium botulinum"
         289..294
         /note="ORF22"
         302..841
         /note="ORF22"
         /codon_start=1
         /transl_table=11
         /protein_id="CAA51308.1"
         /db_xref="GI:509269"
         /db_xref="SPTREMBL:Q38195"
         /translation="MNDLFVAIENLKHDNCHFNFIEMSLKKVYKTSKYNLYDYNN
         DILVHLAKELIENLKNFSELDLRKYSTIKRYCINICKKRDKKIYNSEVYK
         KLDVAVYSLCDNFELDLSILINKKQIILYMKFFEGKDNETAIRLRLSRQSIYK
         IRIISLKKLPYVQLVNI"
         complement(854)..2787
         /gene="ANTP-70"
         854..906
         /note="putative; ORF22"
         complement(854)..906
         /gene="ANTP-70"
         /note="putative"
         complement(916)..2787
         /gene="ANTP-70"

gene
terminator
terminator
CDS

```

```

/codon_start=1
/transl_table=11
/protein_id="CAA51309.1"
/db_xref="GI:509270"
/db_xref="SWISS-PROT:P46085"
/translation="MSLSIKELYITKDSINNVNADGNVYVNRGDSWILSRQNLG
GNISNGCTAIVGDDURIRETATPYPTASFNEEYIKNNVQNVFANFTEASEIPIGPE
FSITAFNSKSLMYLQYTYIRYEIIKVLQNTVERAVLYPSLGVKYSIEFSEQID
KNFYFTSQDKILNEFYIKKIDDTITVKESKNNNINFNSTQITILPRLGLVINK
GDGYMTNDKDLIGTLTIBSSSTGGIIOPRLNNTPLFNTSNPTIFSQETEARLND
AFIQLFNTSTTLFKVEEAPTKNTSMKVNTYEKELINYONGNIDDKAEYLPSL
GKEVSDAPSPQAPVETPDQDGIQTGPENIIVGVINSENFEEISTPIPDQYTY
NIPTIGNACIVLFRVNTIGYKITTNNLEPLIIYEIAGISSNRMMNSNLSDNIK
AIKIITGLARSDAKSLIVSLFKDNYYIRIPISSTTSQIFKRELGNISLDAST
VNILDNLTSGTHYTRQSPDGVNYSIQLTIPGDFNNIASSIFSFRNNQGICTLY
RTESINGYNLITINNSDLLNVEPISLNGATYIFRVKYTELANIYIIFDAYRNS"
complement(3801)..3252
/gene="ANTP-17"
complement(3801)..3241
/gene="ANTP-17"
/codon_start=1
/transl_table=11
/protein_id="CAA51310.1"
/db_xref="GI:516172"
/db_xref="SWISS-PROT:P46083"
/translation="MSSERTPLPNGYKIKSLFSDSLTYTSSGALSFSNTSSLDNOK
WKLEYISSNGRFSNVAPNKYLAIDYDGYIYLSSSNLSLWPKIKAINSYIICLT
SIVNYDIATWIIDNNNIIDQPINLNPNFDINNNSQILKLEK"
complement(3247)..3252
/gene="ANTP-17"
complement(3304)..4176
/gene="ANTP-33"
complement(3304)..4164
/gene="ANTP-33"
/codon_start=1
/transl_table=11
/protein_id="CAA51311.1"
/db_xref="GI:516173"
/db_xref="SPTREMBL:Q38196"
/translation="MSQTNANDLRNNEVFFISPSNNTNKVLDKISQSEVKLWNKLSGA
NOKRLIYDNTKQAYKIKYMDNTSLILTWNAPLSSSVKTDGTNGDNYLLQNTISR
NVIIRNMPNLVLOINIDDTLMVSTQISSNQFESNCIYEALNNRCKLQTLQNS
DRFLSNLANSQIILVQWEDFSKQWIIETNETKSAIYKCOENNRILTQNSNNYV
ETQYSDSLQIYWNVINDLNDASKYILYLNQDTRVLDVYNSQIANGTHIVDSTHGN
TNOQWIINLI"
complement(4171)..4176
/gene="ANTP-33"
4313..8017
/gene="ANTP-139"
4313..8017
/gene="ANTP-139"
4411..4416
/gene="ANTP-139"
4427..8017
/gene="ANTP-139"
/codon_start=1
/transl_table=11
/protein_id="CAA51312.1"
/db_xref="GI:516174"
/db_xref="SPTREMBL:Q38197"
/translation="MDINDLNINSPVDNKNVYIVRAKRTNFFKAFKVPNIWVAPE
RYVEGLDIAEYKLOGGIYDSNLSQDSERENFLQAIILLKRINTISGKQLSLI
STAIPRYGVIGGYSYSPNLTFTGKTPKSNKLSLVTSTIPFPGGYRETNIEBON
KNFYASNVIIFGPGSNIVENVNVIYKKDAENGMTMAEIVFQPLLYKYNKVIDP
AMELTCLKLSLFLYGLPNSLVVYFTRTELONKQPSQNLIDLLISGVDFLEFI
NINFYWFTNSYFPIPSIMFEKYKNTYEIGNNAIGNDIKURLKQKQFQINQDILWNL
NLNYFCQSFNSIIPDRFSNALKHFYKQYTDYTDNYNINFGVNGQNTKLPSNKN
NTIISKPEKVNVLNENNISLAKSIYDGLGTTEDFYSTYKIPYNEEYRFLSDS
NPLNLSIEVDVSIPEIIDINPKDNSNLVFTQITSTTEVTHTALSINYLQAOI
TNNEFTLSDFSQKSVYKSKSLVYSFLDNLSYLETKNDGPDIDTKKYLKKEVF
KNYSPDLNLTQEDSDMCGINEVJAFGKALNLTNSFVEYQDSGALSISKDNL
REPNIEDDISLGLSPDLNKLIEYSKNIYFKKIIFSFLLDQWTEYYSQYFE
LICMAQOSLAQESLYKQIVQNKFTDLSKASIPPDITLKLIIRRETTKTFIDLSNESQIS

```





```

RESULT 8
AB037920      11631 bp      DNA      linear      BCT 24-JAN-2002
LOCUS         Clostridium botulinum genes for ORF-22, HA-70, HA-17, HA-33, NTNHA,
DEFINITION   neurotoxin, complete cds.
ACCESSION    AB037920
VERSION      AB037920.1 GI:6939789
KEYWORDS     neurotoxin; NTNHA; HA-33; HA-17; HA-70; ORF-22.
SOURCE       Clostridium botulinum (strain:D-4947) DNA.
ORGANISM     Clostridium botulinum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE    1
AUTHORS      Kouguchi,H., Watanabe,T., Sagane,Y., Sunagawa,H. and Ohyama,T.
TITLE        In vitro reconstruction of the Clostridium botulinum type D
progenitor toxin
JOURNAL      J. Biol. Chem. 277 (4), 2650-2656 (2002);
MEDLINE      21659747
REFERENCE    2 (bases 1 to 11631)
AUTHORS      Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Takizawa,J.,
TITLE        Kawabe,T., Murakami,F., Muroga,A., Nakatsuka,M. and Ohyama,T.
JOURNAL      Submitted (01-FEB-2000) Tohru Ohyama, Tokyo University of
Agriculture, Faculty of Bioindustry; Yasaka 196, Abashiri, Hokkaido
099-2493, Japan (E-mail:t-oyama@bioindustry.nodai.ac.jp,
Tel:81-152-48-3838(ex.368), Fax:81-152-48-2940)
FEATURES     Location/Qualifiers
             1..11631
             /organism="Clostridium botulinum"
             /strain="D-4947"
             /db_xref="taxon:1491"
             6..545
             /gene="orf-22"
             6..545
             /gene="orf-22"
             /codon_start=1
             /transl_table=11
             /product="ORF-22"
             /protein_id="BAA90656.1"
             /db_xref="GI:6939790"
             /translation="MDLFAENLKHNDHDFIEMLSKKYIEKSKYNYLYDYIN
             DILHMKELIEINLKEIRREYATPYEYTFASFNEZYIRNVQVYFANFTASRPIGFE
             KLEAVNYSLYCEDFEFLDLSILNYKEQIITYMKFECKRDNIEARRLHLRSQSIYK
             IRIKSLKLYPIVMQLVNI"
             complement(621..2492)
             /gene="ha-70"
             complement(621..2492)
             /gene="ha-70"
             /codon_start=1
             /transl_table=11
             /product="HA-70"
             /db_xref="GI:6939791"
             /db_xref="GI:6939791"
             /translation="MSLSIKELYTKDKSINNVNLDAGVYVNRGDSWILSRQNMIG
             GNISNGCTAVGLRIRERATPYEYTFASFNEZYIRNVQVYFANFTASRPIGFE
             FSKTAPSKNGLYMLQTYIRYRIEILKVRNTVIRAVLYVPSLGAKSFNRSQGIID
             ANGYTSDXCKILNEKFIYKRIAEITTAKESSDNTNTLSQTILPYNGLYVINK
             GDNVTRNDKDLIGTLTIENITSGSIIPRLNTRTLPNTSPNITLFSQYEYEARLND
             AFNIGLFTSTLFEVFEAPDNKNIEMKAYNTYERELINTQNGNIADKAEYIPLSL
             KCYESDAPSPQAPVETPVQDGFQIOTGNENILVGINPSENIEIETPIPDYDTY
             NIPTSIQDNNAVLYFTVNTTGVYKINAQNLPLPIVIGESGSDMMNQNTLSNNIK
             ALNVTGDSNNAEYSLIVSLIKKNYIRIPOLSSSTNQLIFKRELGNISDLANST
             VNILDNLNTSGTHYITQSPDVGNVYISQLTIPGDPNNIASSIFSTFRNNQIGIHL
             RUTESINGNITKNYSDDLNVNVEPISLLNGATYIFRVKVTNLNINIFDAIRNS"
             complement(2506..2946)
             /gene="ha-17"
             complement(2506..2946)
             /gene="ha-17"
             /codon_start=1
             /transl_table=11
             /product="HA-17"
             /protein_id="BAA90658.1"
             /db_xref="GI:6939792"

/translation="MSERTFLPNGNYIKSLFSDSLYLYTSSGSLSLFWTSSLDNQK
WKLEYTSSNGFRFSVAPENKYLAYDYGFIYVLSSSNNSLWNPITKAINSYIICITL
SIVNYDYANTIVDNNNTIDQILPLNPFIDNNNSQILKLEK"
complement(3008..3968)
/gene="ha-33"
complement(3008..3968)
/gene="ha-33"
/codon_start=1
/transl_table=11
/product="HA-33"
/protein_id="BAA90659.1"
/db_xref="GI:6939793"
/translation="MSQTNANDLRNNEVEFISPSNNTNKVLDKISQSEVKLNKLSGA
NQWRLYIDNKQAYTIKIMVNDLRLTLDAPLSSVSVAITDNTNNQVWILQDTRISR
NVLIRNMPNLVQLQNTDITLIVTQTSNNQOFFKFSNCIYEALNRNRCKIQTOLNS
DRFLSKNLNSQIIVLQWDFDSSRQKWTIEYNETKSAYTLKCOENRRLYIWNQSNYV
ETVQSTDLSIQYWNINYLNDASKYILYNLQDTRVLDVYNSQTANGTHVIVDSYHGN
TNOQMIINLI"
4129..7719
/gene="ntaha"
4129..7719
/gene="ntaha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAA90660.1"
/db_xref="GI:6939794"
/translation="MDINDDLNINSVDNKNVIVVARKNTFFFAKVPAPNIWAPE
RYXGEPDIAEYKLDGGIYDLSFLQSDSERENFLQAIITLKRINNTISGKQLLSLI
STAIPFYGVGGYSPNITFTGKPKGNKKNLSVSTIPFPFGGRETNIYIESQN
KNFVAGNIVIFGSGNIVNNVICYKNDANGMGTMAEILFQPLLIVKYNFYIDP
AMELTKLISLFLYGLGVKIPVSDLVVYRLTDLNKOFSQNLIDLLISGVDLEFI
NTNPFWTNSYFSNFKMPEYKNIYETIEGNAIGNDIKRLKQKQNSQVDLWNL
NLNFSKEPNSIIPDRFSNALKFYKQYTHDYGDNYINGEANGVQNTKLPLSDKN
TNIISKEPVNLVNNANNISMKSNYIGDAGTTEDFYSTIKIYPNEIYERFNDSD
NFPLNNTSIEVDSIPEDIDINPYKDSDDLFTQITTEEVITHALPVNVLQAI
ITNENFTLSDFSKVWSKDSLVYSLDNLSYLETIKNDGPIPTDKKYLWLKEVF
KNYSFDINLTQEDISQGINEVYFGLNKLINLTNSFEYQNSGPISLISKDNL
SEPNEIDDIIPDLSLGLSFKDLNKLVEYKSNRVYFRKYIFNFQDMWTEYSQYFE
LIMAKOSILQASVQVQIIONKFTDLSKASIPDPTLKLKETTEFTDLSNESQIS
NLVYDNEPLKASICVFVEDIYPKFIYSMEKYINNINIKTREFIQCTNTINDNEKSLI
NSYFTKIDPKINQIAKNFQNSQVQVQKEMLSYQALLFATRGPSNLIEDISGK
NTLQYTESVELYVNGESLYLKPSEIVEYFNSNFTNGLNNTFTICFWRFTKDD
DKTLIKNGNCCGWEIYFDNGLVFEIISDNGNSQSVLSLVNINNNTYIISVDRLL
KQDLIFINDKNVANVSEIQLINLYSTVLSVKNNSYVEELSVDLDPVASEVIR
NYSYLDNSYIRDSKSLLEYKNYOLNYVPETSLYEVDNKNKSLKNTDGINI
PSVKFKLINTDESKGVQKWDECIQVSDGTERYLDISPENNRIOLVSSKDNNAKITV
NTDLPRDPCITFSYNDKYFSLSLRDGDYNMIMCNDNKKVPKGHLWLKAS"
7734..11591
/gene="nt"
7734..11591
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BAA90661.1"
/db_xref="GI:6939795"
/translation="MTWPKDFNYSQVNDNDILYRIPONKLIITTPVAFNMITONIW
VIPERFSDNPSLSRPREPTKIYQYDPSYLSLTDQKDTLKGILKLFKRINERDI
GKILINLYVSGSPMGDSPTDFTRHTTNIAVEKFGNSWKVNTIITPSVLIFG
PLPNILDTASLTLOGQCSNPSPGFGT-SILKVAPEFLTFSDVNSQSAVLGKSI
FCMDPVIALMHELTGS-HOLYGINIPSDKRIPOVSEGFSDGPNVQFEELYTFGGS
DVELIPIQLRQLRQKALGHYKIDIAKRLANNIKTIPSSMSNIDKVKKIFSEKYNFK
DNTINGVNIIDKNSLSLDTNVSEVYSSQNVKNRPHYFSKHYLPVFNILDDNI
XTINGNLTGKFNISNGONIERPAKQSSSEVDLFTKVLRLRNRSDSTC
IQVKNNTLPYADKDSIEFISQITDENTVNSDNFSDDESIDKARITNPEAV
DPLLPNTYMEPLNVPGEVEFYDDITKDVDLNSYIYLEAQKLSNNENITLTSVEE
ALGYSNKIYTFPLSLAEKVNKGQAGLFNWNAVEVDEFTNIMKKDTLDKISDVSAI
IPYIGPALNTGNSALRGNFKQATAGVAFLEGFPEFTIPALGVFTFYSSIQEREKI
IKTENCLEQVRKWKDSYOMVSNMLRITTRFNHISQMDYLSYQADAIKAKIDL
EYKYSQSDRENTKSNQVENKSLDKISEAMNNINKFIRECSVYLFKNMLPKVIDE
LNKFDLKTTELINLIDSHNIIIVGEVDRLKAKVNSNFENFTFNIFSTINNSLAKDM
INEYFNISNDSKILSLQNKNTLMDTSGYNAEVRVEGNVOLNPFIFFDFKLGSSGDDR

```

GVIVTQNEIVNAMYSESISFWIRINKWYNLPGYTIIDSVKNNWSGSIIGSNF  
LFTLKQNSQEDINFSYISKNAAGYKWFVITTTNMGNMIIYIINGKIDIKV  
KELTGINFSTIIFOMNKIDITSDSDNIWMIDFYIFAKELDDKIDILNLSL  
QYTNVVDYWGNDLRYDKEYIMLVNVMNMYKSGKNGIVENTRNKNDNFNGYKII  
KRIGTNDTRVRGENVLYENTIDNKQYSLNGKYSRLGTDLPGLNLDPMDEIR  
KYGSIIFOPNTEDYASQIFLSNATNRLGILISGSYFKLGDDIWFNHEYLIPVI  
KIEHYASLLESTSHWVFVASE"  
BASE COUNT 4493 a 1254 c 1657 g 4227 t  
ORIGIN

Query Match 32.8%; Score 449.8; DB 1; Length 11631;  
Best Local Similarity 59.38; Pred No. 1.3e-59;  
Matches 807; Conservative 0; Mismatches 542; Indels 12; Gaps 2;  
QY 13 ACCATCCCATCAACATCTTCCTCTACACCAACATCCCTGTTGAAGSACATCATCAAC 72  
DB 10248 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTATTAAGAAGATATGATTAAT 10307  
QY 73 GAGTACTTCAACAATCAACAGACTCCAGATCTGCTCCCTGCGAGACCGCTAAGACACC 132  
DB 10308 GAATATTTCAATAGTATTAATGATTCAAAAATTTTGAGCTTTACAAAAATAAAAAATACT 10367  
QY 133 TTGGTGCACACTCCGGTTACAACGCCGAGTCTCCGAGAGGGTGACGTCCAGCTGAAC 192  
DB 10368 TTGATGGATACATCAGGATATAACSCAGAGTGAAGTGAAGGCAATGTTTCAGCTTAAT 10427  
QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGAGTAAAGTCAATC 252  
DB 10428 CCAATATTTCCATTTGACTTTAAATAGTAGTTCAGGGGATGATAGAGTAAAGTTATA 10487  
QY 253 GTCAACCCAGAGAGAGACATCGTCTACAACATCCCATGATGACAGTCTCTCTCCATCTCCTC 312  
DB 10488 GTAAACCCAGAGTAAATATTGTATATATGCTATGATGAAGTTTTAGTATTAGTTTT 10547  
QY 313 TGGATCAGNATCAACAAGTGGTCTCCAACTTCCAGGTTCACACATCATCGACTCCGTC 372  
DB 10548 TGGATAGGATAAATAATGGTGAAGTAAATTTACCTCGATATACTATAATGATAGTGT 10607  
QY 373 AAGAACAACCTCCGGTGGTTCATCGGTATCATCTCCAACTTCCCTGGTTCACCCCTGAAG 432  
DB 10608 AAAAATAACTCAGGTGGAGTATAGTATTATAGTAAATTTTGTAGTCTTTTAAATA 10667  
QY 433 CAGACAGGAGCTCCGAGAGTCCCATCAACTTCTCTAGCATCTCCACACAGCTCTCT 492  
DB 10668 CAAATGAAATAGTAGACAAGATATAAATCTTAAATGATGATATCAAAAGATGCTGG 10727  
QY 493 GGTTCACAACAAGTGGTCTTCCTGCTCACCCTCCACACACATGATGGGTAAACATGAAGATC 552  
DB 10728 GGNATATAATAGTGTTTTGTACTATTACTACCAATATGATGGGAATATGATGATT 10787  
QY 553 TACATCAACGGTGAAGCTGATGACACCAATCAAGGTCAAGGTGACCGGTATCACTTC 612  
DB 10788 TATATAAATGGAATTAATAGATATACTATAAAGTTAAAGAGTTAACTGGAATTAATTT 10847  
QY 613 TCAAGACCATCACCTTCGAGATCAACAAGATCCAGACACCGGTCTGATCACCTCCGAC 672  
DB 10848 AGCAAACTTATACATTTTCATGATTAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10907  
QY 673 TCGACCAACATCAACATGGGATCGGTGACCTTCTACATCTTCCTGCGCAGAGGTGGAGGT 732  
DB 10908 TCTGATACATCAATATGGGATAGGATTTTATATCTTTTGTGTAAGGATTAAGATGAT 10967  
QY 733 AAGGACATCAACATCTGTTCACTTCTGAGTACACCAACCGCTCGTCAAGGACTACTGG 792  
DB 10968 AAGATATTAATATATTTTATAGCTTGCATATGATGATGATGATGATGATGATGATGATG 11027  
QY 793 GGTAAAGCTGAGATACACAAGGAGTACATGTTGTTCAACATCGACTACTTCTGAAACAGA 852  
DB 11028 GGAATAGATTTAAGATATGATAAAGATATTACATGATTAACTGTAATTAATGATGATG 11087  
QY 853 TACATCTAGCCCACTCCAGACAGATCTCTCTACACACCGCTTACACACAGGACTTC 912  
DB 11088 TATATGCTCTAAAAAAGGCAATGGAATGTTTATACACCGCTAAAAATAATAGTCTTC 11147

QY 913 AACGAGGTTTAAAGATCATCATCAAGCGTATCAGAGTAAACCAACAGCACAGATC 972  
DB 11148 AATGAAGGATATAAATTTATAAATAAAGAAATAGAGGAAATAAATGATGATACTAGATG 11207  
QY 973 AGAGTGTGTGACATCTCTGCTAGTCTGACATGATCAACACACAGGCTTACACCTGTTCC 1032  
DB 11208 CGAGGAGAAATGATATATATTTTAACTACAAATGATACAAACAATATAGTTAGGT 11267  
QY 1033 ATGAAGAAGAGAGACATGATACGCGCAACACCATCTCCACCGAGGACATCTACGCGATCGGT 1092  
DB 11268 ATGATATAACCTTCTA-----GAAATCTAGGAGTATTTAGTTCCACTAGGTGCA 11318  
QY 1093 CTGCGTGAAGAGACCAAGGACATCAACGACACATCATCTTCAGATCCAGCAACAAGAAC 1152  
DB 11319 TTGGATCAACCAATGGATGAGATACGTAATATGTTGTTTATATACACCAATGCAAT 11378  
QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212  
DB 11379 ACTTTTGATTAATGATGATCAATATTTTGTCAAGTATGCAACAACAATAGGCTT 11438  
QY 1213 GGTATCTTTCATCGGTACCTACAGATTCGGTCTGGTGGTGAC---TGTACAGACAC 1269  
DB 11439 GGAATPACTATCAATTTGGTATGTTTCAAGTTCGAGATGACTATTTGTTTAAATCAC 11498  
QY 1270 AACTACTTGGTTCCAACGTCAACGAGGTAACTACGCTTCTGCTGGAGTCCACTTCC 1329  
DB 11499 GAATATTTAATCTGTTATATAAATAGAGCATTTATGCTTATATAGATCAACATCA 11558  
QY 1330 ACCCACTGGGATGTCCTCCAGTCTCCGAGTAAATAGGAAT 1370  
DB 11559 ACTCATTTGGTGGTGGTACCTGCAAGTGAATATAAATAGT 11599  
RESULT 9  
CLOTNCO 4049 bp DNA linear BCT 18-DEC-2001  
LOCUS Clostridium botulinum type D gene for neurotoxin consisting of  
DEFINITION botulinum neurotoxin D and Cl, complete cds.  
ACCESSION D38442  
VERSION D38442.1 GI:1374775  
KEYWORDS neurotoxin consisting of botulinum neurotoxin D and Cl; neurotoxin.  
SOURCE Clostridium botulinum D (strain:South African).  
ORGANISM Clostridium botulinum D  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
REFERENCE 1 Moriishi, K., Koura, M., Abe, N., Fujii, N., Fujinaga, Y., Inoue, K. and Ogumad, K.  
Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms Biochim. Biophys. Acta 1307 (2), 123-126 (1996)  
AUTHORS 2 (bases 1 to 4049)  
Moriishi, K.  
TITLE Direct Submission  
JOURNAL Submitted (25-SEP-1994) Kohji Moriishi, Osaka University, Research Institute for Microbial Diseases, Research Center for Emerging Infectious Diseases; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kohji@biken.osaka-u.ac.jp, Tel:81-6-6879-8343, Fax:81-6-6879-8269)  
AUTHORS Moriishi, K.  
TITLE Direct Submission  
JOURNAL  
FEATURES  
source 1. 4049  
/organism="Clostridium botulinum D"  
/strain="South African"  
/db\_xref="taxon:56829"  
163..4020  
/codon\_start=1  
/transl\_table=11  
/product="neurotoxin consisting of botulinum neurotoxin D and Cl"  
/protein\_id="BA007477.1"  
/db\_xref="GI:1374776"

BASE COUNT	1621 a	438 c	630 g	1360 t	
ORIGIN					
Query Match	32.7%	Score 448.2	DB 1	Length 4049	
Best Local Similarity	59.2%	Pred. No. 2.6e-59			
Matches	806	Conservative 0	Mismatches 543	Indels 12	Gaps 2
QY	13	ACATCCCATCAACATCTTCTCTACACACAACTCCCTGTGAGGACATCATCAAC	72		
DB	2677	ACAAATACCCCTTAATATTTTTCATATACTAAATATCTTTATTAAGATATGATTAAT	2736		
QY	73	GAGTACTTCAACAAACATCAACGACTCCAAGATCCTGCTCGACAGACCGTAAGAACACC	132		
DB	2737	GATATTTCAATAGTATTAATGATTCAAAATTTTGAGCTTACAAATPAAAAAATAACT	2796		
QY	133	TTGGTCGACACCTCCGGTTTCAACCGCGAGGTCTCCGAGGAGGGTGAAGTCACGCTGAAC	192		
DB	2797	TTGATGGATACATCAGGATATAACGCCAGAAGTGAAGTGAAGGCAATGTCAGCTTAAT	2856		
QY	193	CCAACTCTCCCAATTCGACTTCAAGCTGGTCTCCGGTGAGACAGAGGTAAAGTCATC	252		
DB	2857	CCAAATTTCCCAATTTGACTTTAAATTAGTAGTTCAGGGATGATAGAGGTAATAATTATA	2916		
QY	253	GTCAACCCGAACGAGAAACATCGTCTCAAACTCCATGTACGAGTCTTCCCATCTCCCTC	312		
DB	2917	GTAACCCGAATGAAATATTGTATATAATCCATGTATGAAGTTTGTAGTATTAGTTT	2976		
QY	313	TGATCAGAAATCAACAGTGGGTCTCCAATCTGCGAGGTTTACACCATATCGACTCCGTC	372		
DB	2977	TGGATTAGGATAAATAATGGGTAAAGTAATTAACCTGGATACTATATAATTGATAGTGT	3036		
QY	373	AGAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACCTTCTGTGCTTCCACCTGAAG	432		
DB	3037	AAAATACTCAGGTGAGATATAGGTATTATTAGTAATTTTTTTAGTGTACTTTAAAA	3096		
QY	433	CAGAACGAGGACTCCGACAGCTCCCAACTTCTCTACGACATCTCCAACACGCTCCT	492		
DB	3097	CRAAATGAAAATAGTGAACAAGATATAAACTTTAGTTATGATATATCAAGAATGCTGCG	3156		
QY	493	GGTTACACAAGTGGTTCTTCGTACCGGTCAACACACATGATGGTAAACATGAAGATC	552		
DB	3157	GGATATAATAATGGTTTTTTTGTAACTATTACTACCAATAATGATGGGAATAATGATGATT	3216		
QY	553	TACATCAACGGTAAAGCTGATCGACACCATCAAGGTCAAGGATGTGACCGGTATCAACTTC	612		
DB	3217	TATATAAATGGAAATTAATAGATACTATAAAGTTAAAGAGTTACTGGAATTAATTTT	3276		
QY	613	TCCAAGACATCACCTTCGAGATCAACAGATCCAGACACCGGCTGTGATCATCCTCGAC	672		
DB	3277	AGCAAACTATAAACAATTTCAATGAATAAATAATCCAAATTAACCTGGCTTAATTACCTCAGAT	3336		

```

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabsi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkenko,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivas,M., Rojas,A., Rojubokan,I., Roife,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usamai,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 151705)
Worley,K.C.
Direct Submission
Submitted (11-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 151705)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20162890.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVFG
Center clone name: CH230-320C16
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 121907 bases at least Q40
Consensus quality: 127236 bases at least Q30
Consensus quality: 130792 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1 1173: contig of 1173 bp in length

```

```

* 1174 1273: gap of unknown length
* 1274 2448: contig of 1175 bp in length
* 2449 2548: gap of unknown length
* 2549 3932: contig of 1384 bp in length
* 3933 4032: gap of unknown length
* 4033 5391: contig of 1159 bp in length
* 5392 5291: gap of unknown length
* 5292 6511: contig of 1219 bp in length
* 6511 6611: gap of unknown length
* 6611 7729: contig of 1119 bp in length
* 7730 7829: gap of unknown length
* 7830 9652: contig of 1823 bp in length
* 9653 9753: gap of unknown length
* 9753 12017: contig of 2264 bp in length
* 12017 13885: contig of 1769 bp in length
* 13886 13985: gap of unknown length
* 13986 15659: contig of 1684 bp in length
* 15660 15759: gap of unknown length
* 15760 17357: contig of 1598 bp in length
* 17358 17467: gap of unknown length
* 17468 19346: contig of 1879 bp in length
* 19347 19446: gap of unknown length
* 19447 20660: contig of 1214 bp in length
* 20661 20760: gap of unknown length
* 20761 23066: contig of 2306 bp in length
* 23067 23166: gap of unknown length
* 23167 25620: contig of 2454 bp in length
* 25621 25740: gap of unknown length
* 25741 28298: contig of 2478 bp in length
* 28299 31368: gap of unknown length
* 31369 31468: gap of unknown length
* 31469 35026: contig of 3558 bp in length
* 35027 35126: gap of unknown length
* 35127 38571: contig of 3444 bp in length
* 38572 38670: gap of unknown length
* 38671 42549: contig of 3879 bp in length
* 42550 42650: gap of unknown length
* 42651 45722: contig of 3073 bp in length
* 45723 45822: gap of unknown length
* 45823 49824: contig of 4002 bp in length
* 49825 49924: gap of unknown length
* 49925 53901: contig of 3977 bp in length
* 53902 54001: gap of unknown length
* 54002 58137: contig of 4136 bp in length
* 58138 58237: gap of unknown length
* 58238 63191: contig of 4954 bp in length
* 63192 63291: gap of unknown length
* 63292 69224: contig of 5933 bp in length
* 69225 69324: gap of unknown length
* 69325 75399: contig of 6075 bp in length
* 75400 75499: gap of unknown length
* 75500 81756: contig of 6257 bp in length
* 81757 81856: gap of unknown length
* 81857 88321: contig of 6465 bp in length
* 88322 88421: gap of unknown length
* 88422 98795: contig of 10374 bp in length
* 98796 98895: gap of unknown length
* 98896 112987: contig of 14092 bp in length
* 112988 113087: gap of unknown length
* 113088 127886: contig of 14799 bp in length
* 127887 127986: gap of unknown length
* 127987 151705: contig of 23719 bp in length.
Location/Qualifiers
1. 151705
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-320C16"
BASE COUNT 42411 a 30154 c 29282 g 44163 t 5695 others
ORIGIN
Query Match 13.8%; Score 189; DB 2; Length 151705;

```



```

----- Project Information
Center project name: Gfr1
Center clone name: CH240-358N16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 120888 bases at least Q40
Consensus quality: 127504 bases at least Q30
Consensus quality: 132322 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 61 contigs, the true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 1019: contig of 1019 bp in length
  * 1020 1119: gap of unknown length
  * 1120 2779: contig of 1660 bp in length
  * 2780 2879: gap of unknown length
  * 2880 4496: contig of 1617 bp in length
  * 4497 4596: gap of unknown length
  * 4597 5621: contig of 1025 bp in length
  * 5622 5721: gap of unknown length
  * 5722 6889: contig of 1168 bp in length
  * 6890 6989: gap of unknown length
  * 6990 8124: contig of 1135 bp in length
  * 8125 8224: gap of unknown length
  * 8225 9804: contig of 1580 bp in length
  * 9805 9904: gap of unknown length
  * 9905 11392: contig of 1488 bp in length
  * 11393 11492: gap of unknown length
  * 11493 12676: contig of 1184 bp in length
  * 12677 12776: gap of unknown length
  * 12777 13861: contig of 1085 bp in length
  * 13862 13961: gap of unknown length
  * 13962 15707: contig of 1746 bp in length
  * 15708 15807: gap of unknown length
  * 15808 17582: contig of 1775 bp in length
  * 17583 17682: gap of unknown length
  * 17683 19005: contig of 1323 bp in length
  * 19006 19105: gap of unknown length
  * 19106 21071: contig of 1966 bp in length
  * 21072 21171: gap of unknown length
  * 21172 22855: contig of 1684 bp in length
  * 22856 22955: gap of unknown length
  * 22956 24423: contig of 1468 bp in length
  * 24424 24523: gap of unknown length
  * 24524 26272: contig of 1749 bp in length
  * 26273 26372: gap of unknown length
  * 26373 28549: contig of 2177 bp in length
  * 28550 28649: gap of unknown length
  * 28650 30139: contig of 1490 bp in length
  * 30140 30239: gap of unknown length
  * 30240 31817: contig of 1578 bp in length
  * 31818 31917: gap of unknown length
  * 31918 33762: contig of 1845 bp in length
  * 33763 33862: gap of unknown length
  * 33863 36122: contig of 2260 bp in length
  * 36123 36222: gap of unknown length
  * 36223 38477: contig of 2255 bp in length
  * 38478 38577: gap of unknown length
  * 38578 40590: contig of 2013 bp in length
  * 40591 40690: gap of unknown length
  * 40691 43154: contig of 2464 bp in length
  * 43155 43254: gap of unknown length
  * 43255 45512: contig of 2258 bp in length
  * 45513 45612: gap of unknown length

```

```

* 45613 47325: contig of 1713 bp in length
* 47326 47425: gap of unknown length
* 47426 49849: contig of 2424 bp in length
* 49850 49849: gap of unknown length
* 49851 51626: contig of 1677 bp in length
* 51627 51726: gap of unknown length
* 51727 52957: contig of 1131 bp in length
* 52958 52957: gap of unknown length
* 52959 56168: contig of 3211 bp in length
* 56169 56268: gap of unknown length
* 56270 59010: contig of 2742 bp in length
* 59011 59110: gap of unknown length
* 59111 61204: contig of 2094 bp in length
* 61205 61304: gap of unknown length
* 61305 63328: contig of 2024 bp in length
* 63329 63428: gap of unknown length
* 63429 66088: contig of 2660 bp in length
* 66089 66188: gap of unknown length
* 66189 67828: contig of 1640 bp in length
* 67829 67928: gap of unknown length
* 67929 70649: contig of 2721 bp in length
* 70650 70749: gap of unknown length
* 70750 72883: contig of 2134 bp in length
* 72884 72983: gap of unknown length
* 72984 75071: contig of 2088 bp in length
* 75072 75171: gap of unknown length
* 75172 78234: contig of 3063 bp in length
* 78235 78334: gap of unknown length
* 78335 80048: contig of 1714 bp in length
* 80049 80148: gap of unknown length
* 80149 83263: contig of 3115 bp in length
* 83264 83633: gap of unknown length
* 83634 85719: contig of 2356 bp in length
* 85720 85819: gap of unknown length
* 85820 87592: contig of 1773 bp in length
* 87593 87692: gap of unknown length
* 87693 91127: contig of 3435 bp in length
* 91128 91227: gap of unknown length
* 91229 94240: contig of 3013 bp in length
* 94241 94340: gap of unknown length
* 94341 98200: contig of 4480 bp in length
* 98201 98920: gap of unknown length
* 98921 104013: contig of 5093 bp in length
* 104014 104113: gap of unknown length
* 104114 107600: contig of 3487 bp in length
* 107601 107700: gap of unknown length
* 107701 111658: contig of 3958 bp in length
* 111659 111758: gap of unknown length
* 111759 115121: contig of 3363 bp in length
* 115122 115221: gap of unknown length
* 115222 120413: contig of 5192 bp in length
* 120414 120513: gap of unknown length
* 120514 125376: contig of 4862 bp in length
* 125377 125475: gap of unknown length
* 125476 129676: contig of 4201 bp in length

```

Query Match 13.7%; Score 188.4; DB 2; Length 176351;

Best Local Similarity 49.6%; Pred. No. 9.3e-20;

Matches 493; Conservative 0; Mismatches 491; Indels 0; Gaps 0;

```

QY 236 ACAGAGTAAGTTCCTGTCACCCAGACGAGACATCGTCTACAACTCCATGTACGAGT 295
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 52708 ACAGACACAGCAGCAGCAACACACGACGACCAACACACACACACACACACGCA 52649
QY 296 CCCTTCCTCATCTCTTCTGATCGAATCAACAAGTGGGTCTCCAACTTGCCAGGTTACA 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 52648 ACACACACACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACA 52589
QY 356 CCATCATCGACTCCGTCAGACAACTCCGGTGGTTCCTCGGTATCATCTCCAACTTCC 415
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 52588 GCAGCAACACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACA 52529
QY 416 TGGTCTTCACCCCTGAAGCAGACGAGGACTCCGAGCAGTCCATCATCTCTCTAGGACA 475

```







\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1195: contig of 1195 bp in length
* 1196 1295: gap of unknown length
* 1296 2301: contig of 1006 bp in length
* 2302 2401: gap of unknown length
* 2402 3597: contig of 1196 bp in length
* 3598 3697: gap of unknown length
* 3698 4705: contig of 1008 bp in length
* 4706 4805: gap of unknown length
* 4806 6053: contig of 1248 bp in length
* 6054 6153: gap of unknown length
* 6154 7709: contig of 1556 bp in length
* 7710 7809: gap of unknown length
* 7810 9075: contig of 1266 bp in length
* 9076 9175: gap of unknown length
* 9176 10387: contig of 1212 bp in length
* 10388 10487: gap of unknown length
* 10488 12171: contig of 1684 bp in length
* 12172 12271: gap of unknown length
* 12272 13923: contig of 1852 bp in length
* 13924 14233: gap of unknown length
* 14234 15353: contig of 1230 bp in length
* 15354 15353: gap of unknown length
* 15354 17056: contig of 1703 bp in length
* 17057 17156: gap of unknown length
* 17157 18630: contig of 1474 bp in length
* 18631 18730: gap of unknown length
* 18731 20564: contig of 1834 bp in length
* 20565 20664: gap of unknown length
* 20665 22920: contig of 2256 bp in length
* 22921 23020: gap of unknown length
* 23021 24042: contig of 1022 bp in length
* 24043 24142: gap of unknown length
* 24143 25407: contig of 1265 bp in length
* 25408 25507: gap of unknown length
* 25508 27326: contig of 1819 bp in length
* 27327 27426: gap of unknown length
* 27427 28625: contig of 1399 bp in length
* 28626 28925: gap of unknown length
* 28926 30295: contig of 1370 bp in length
* 30296 30395: gap of unknown length
* 30396 32901: contig of 2506 bp in length
* 32902 33001: gap of unknown length
* 33002 35383: contig of 2382 bp in length
* 35384 35483: gap of unknown length
* 35484 38276: contig of 2793 bp in length
* 38277 38376: gap of unknown length
* 38377 39868: contig of 1492 bp in length
* 39869 39968: gap of unknown length
* 39969 41970: contig of 2001 bp in length
* 41970 42069: gap of unknown length
* 42070 44382: contig of 2313 bp in length
* 44383 44482: gap of unknown length
* 44483 47592: contig of 3110 bp in length
* 47593 47592: gap of unknown length
* 47593 52151: contig of 4459 bp in length
* 52152 52451: gap of unknown length
* 52452 56760: contig of 4509 bp in length
* 56761 56860: gap of unknown length
* 56861 61177: contig of 4317 bp in length
* 61178 61277: gap of unknown length
* 61278 66148: contig of 4871 bp in length
* 66149 66248: gap of unknown length
* 66249 71886: contig of 5638 bp in length
* 71887 71986: gap of unknown length
* 71987 78079: contig of 6093 bp in length
* 78080 78179: gap of unknown length
* 78180 81154: contig of 2975 bp in length
* 81155 81254: gap of unknown length

```

```

* 81255 88890: contig of 7636 bp in length
* 88891 88990: gap of unknown length
* 88991 97276: contig of 8286 bp in length
* 97277 97376: gap of unknown length
* 97377 107376: contig of 10000 bp in length
* 107377 107476: gap of unknown length
* 107477 116754: contig of 9278 bp in length
* 116755 116854: gap of unknown length
* 116855 127329: contig of 10475 bp in length
* 127330 127429: gap of unknown length
* 127430 136746: contig of 9317 bp in length
* 136747 136846: gap of unknown length
* 136847 153947: contig of 17101 bp in length
* 153948 154047: gap of unknown length
* 154048 169163: contig of 15116 bp in length.
FEATURES
  source
    1. 169163
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-261G15"
BASE COUNT 41981 a 37857 c 38661 g 43240 t 7424 others
ORIGIN
Query Match 13.3%; Score 183; DB 2; Length 169163;
Best Local Similarity 50.6%; Pred. No. 6.3e-19;
Matches 469; Conservative 0; Mismatches 455; Indels 3; Gaps 1;
QY 233 AGACAGAGGTAAAGTTCATCGTCACCCAGACGAGCAACATCGTTACAACTCCCATGTACG 292
DB 35132 AAGCCAAACCAACACCAACACAGCAGCAACAAACAAACAGCAACAGCAGCA 35073
QY 293 AGTCCTTCCTTCCTCTCTCTGATCAGATCAACAAAGTGGTCTCCAACTTGGCAGTT 352
DB 35072 ACAACAGCAACAGCAACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAA 35013
QY 353 ACACCATCATCGACTCCGTCACAGCAACAACTCCGGTTGGTTCATCGGTATCTCCCACT 412
DB 35012 ACAGCAACAAACAGCAGCAGCAACATACACAGCAACAAACAAACATCAGCAACA 34953
QY 413 TCCTGTCTTCACCTCGGAGAGCAGAGGAGTCCGAGAGTCCATCACTTCCTCTACG 472
DB 34952 ACAACAAACAGCAACAGCAACAAACAGCAACAGCAACAAACAGCAACAGCAACA 34893
QY 473 ACATCTCCAAACAGCTCTCTGTTTACAAAGTGGTCTTCGTCACCGTCCACCAACA 532
DB 34892 ACACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAAACAAACAGCAACACA 34833
QY 533 TGATGGTTAACATGAGATCTACATCAACGGTAACTGATCGACACCATCAAGGTCAAG 592
DB 34832 GCAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAAACA 34773
QY 593 AGTTGACCGGTATCACTTCTCAAGACCATCACTTCGAGATCAACAGATCCAGACA 652
DB 34772 ACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACA 34713
QY 653 CCGGTCTGATCACTCCGACTCCGACATCAACATCAACATGGTGGTGGTCTTCTTCTTCT 712
DB 34712 ACAGCAACAAACAGCAACAAACAGCAACAAACAGCAACAGCAACAGCAACAAACA 34653
QY 713 TCGCCAAGGAGTTGGCGTAAAGGACATCAACATCTCTTCACTCTCTTGA---GTACA 769
DB 34652 GCAACAAACAGCAGCAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACA 34593
QY 770 CCAACCTCGTCAAGGACTACTGGGGTAAAGCTGAGATACAGCAACAGGAGTACTACATGG 829
DB 34592 ACAACAAACAGCAGCAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAACA 34533
QY 830 TCAACATCGACTACTTGACACAGATACATGTAGCGCAACTCCAGACAGATCTCTTCTTACA 889
DB 34532 ACAACAAACAGCAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACA 34473
QY 890 CCAGAGGTAAACAAACAGCACTTCAACGAGGGTTACAAGATCATCATCAAGCGTTCAGAG 949

```

**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
  
**COMMENT**

Direct Submission  
 Unpublished  
 2 (bases 1 to 163034)  
 Worley,K.C.  
 Direct Submission  
 Submitted (15-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 163034)  
 Worley,K.C.  
 Direct Submission  
 Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 11, 2002 this sequence version replaced gi:17973419.  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GHWE  
 Center clone name: CH230-103G20  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 79006 bases at least Q40  
 Consensus quality: 84121 bases at least Q30  
 Consensus quality: 88465 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 82 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1382: contig of 1382 bp in length  
 \* 1383 1482: gap of unknown length  
 \* 1483 2616: contig of 1134 bp in length  
 \* 2617 2716: gap of unknown length  
 \* 2717 4049: contig of 1333 bp in length  
 \* 4050 4149: gap of unknown length  
 \* 4150 5644: contig of 1495 bp in length  
 \* 5645 5744: gap of unknown length  
 \* 5745 5913: contig of 1169 bp in length  
 \* 5914 7013: gap of unknown length  
 \* 7014 8388: contig of 1375 bp in length  
 \* 8389 8489: gap of unknown length  
 \* 8489 10138: contig of 1650 bp in length  
 \* 10139 10238: gap of unknown length  
 \* 10239 11735: contig of 1496 bp in length  
 \* 11735 11834: gap of unknown length  
 \* 11835 13287: contig of 1453 bp in length  
 \* 13288 13387: gap of unknown length  
 \* 13388 14709: contig of 1322 bp in length  
 \* 14710 14809: gap of unknown length  
 \* 14810 15861: contig of 1052 bp in length  
 \* 15862 15961: gap of unknown length  
 \* 15962 17518: contig of 1557 bp in length  
 \* 17519 17618: gap of unknown length  
 \* 17619 19300: contig of 1682 bp in length  
 \* 19301 19400: gap of unknown length  
 \* 19401 20729: contig of 1329 bp in length  
 \* 20730 20829: gap of unknown length  
 \* 20830 21929: contig of 1100 bp in length  
 \* 21930 22029: gap of unknown length  
 \* 22030 23415: contig of 1386 bp in length  
 \* 23416 23515: gap of unknown length



ACCESSION AC096212  
 VERSION 4  
 KEYWORDS GI-21723349  
 SOURCE HTGS\_PHASE1  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 165337)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C., Ausbrooks, S.L., Amaral, T., Bimaye, K., Blankenbuehler, K., Bonnini, D., Bouck, J., Bowie, S., Briveau, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escott, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Hernandez, O., Hodgson, A., Hoques, M., Hollaway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, F., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Metker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweso, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 165337)  
 AUTHOR Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 165337)  
 AUTHOR Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Jul 10, 2002 this sequence version replaced gi:17943903.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: 35RC  
 Center clone name: CH230-23N22  
 ----- Summary Statistics

Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 96557 bases at least Q40  
 Consensus quality: 103304 bases at least Q30  
 Consensus quality: 107916 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 68 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1019: contig of 1019 bp in length  
 \* 1020 1119: gap of unknown length  
 \* 1120 2182: contig of 1063 bp in length  
 \* 2183 2282: gap of unknown length  
 \* 2283 3490: contig of 1208 bp in length  
 \* 3491 3590: gap of unknown length  
 \* 3591 4592: contig of 1002 bp in length  
 \* 4593 4592: gap of unknown length  
 \* 4593 5977: contig of 1285 bp in length  
 \* 5978 6077: gap of unknown length  
 \* 6078 7186: contig of 1109 bp in length  
 \* 7187 7286: gap of unknown length  
 \* 7287 8413: contig of 1127 bp in length  
 \* 8414 9632: contig of 1169 bp in length  
 \* 9633 9632: gap of unknown length  
 \* 9633 10917: contig of 1135 bp in length  
 \* 10918 11017: gap of unknown length  
 \* 11018 12031: contig of 1034 bp in length  
 \* 12032 12151: gap of unknown length  
 \* 12152 13374: contig of 1223 bp in length  
 \* 13375 13474: gap of unknown length  
 \* 13475 14588: contig of 1114 bp in length  
 \* 14589 14688: gap of unknown length  
 \* 14689 16062: contig of 1374 bp in length  
 \* 16063 16162: gap of unknown length  
 \* 16163 17290: contig of 1128 bp in length  
 \* 17291 18879: contig of 1489 bp in length  
 \* 18880 18979: gap of unknown length  
 \* 18980 20557: contig of 1578 bp in length  
 \* 20558 21994: gap of unknown length  
 \* 21995 22094: gap of unknown length  
 \* 22095 23618: contig of 1524 bp in length  
 \* 23619 23718: gap of unknown length  
 \* 23719 24845: contig of 1127 bp in length  
 \* 24846 24945: gap of unknown length  
 \* 24946 26434: contig of 1488 bp in length  
 \* 26435 26533: gap of unknown length  
 \* 26534 28392: contig of 1859 bp in length  
 \* 28393 28492: gap of unknown length  
 \* 28493 30382: contig of 1890 bp in length  
 \* 30383 30482: gap of unknown length  
 \* 30483 32223: contig of 1741 bp in length  
 \* 32224 32323: gap of unknown length  
 \* 32324 33841: contig of 1518 bp in length  
 \* 33842 33941: gap of unknown length  
 \* 33942 35473: contig of 1532 bp in length  
 \* 35474 35574: gap of unknown length  
 \* 35575 36870: contig of 1297 bp in length  
 \* 36871 36970: gap of unknown length  
 \* 36971 38600: contig of 1630 bp in length  
 \* 38601 38700: gap of unknown length  
 \* 38701 40028: contig of 1328 bp in length  
 \* 40029 40128: gap of unknown length

```

* 40129 41493: contig of 1365 bp in length
* 41593: gap of unknown length
* 41594 42011: contig of 1318 bp in length
* 42912 43011: gap of unknown length
* 44284 44283: contig of 1272 bp in length
* 44384 44383: gap of unknown length
* 44384 46493: contig of 2110 bp in length
* 46494 46593: gap of unknown length
* 46594 48001: contig of 1408 bp in length
* 48002 48101: gap of unknown length
* 48102 50018: contig of 1917 bp in length
* 50019 50118: gap of unknown length
* 50119 51837: contig of 1719 bp in length
* 51838 51937: gap of unknown length
* 51938 54876: contig of 2939 bp in length
* 54877 54976: gap of unknown length
* 54977 56610: contig of 1634 bp in length
* 56611 56710: gap of unknown length
* 56711 58795: contig of 2085 bp in length
* 58796 58895: gap of unknown length
* 58896 60962: contig of 2067 bp in length
* 60963 61062: gap of unknown length
* 61063 63317: contig of 2255 bp in length
* 63318 63417: gap of unknown length
* 63418 64835: contig of 1418 bp in length
* 64836 64935: gap of unknown length
* 64936 66869: contig of 1934 bp in length
* 66870 66969: gap of unknown length
* 66970 69172: contig of 2203 bp in length
* 69173 69272: gap of unknown length
* 69273 71451: contig of 2179 bp in length
* 71452 71551: gap of unknown length
* 71552 73445: contig of 1794 bp in length
* 73446 73445: gap of unknown length
* 73446 76208: contig of 2663 bp in length
* 76109 76208: gap of unknown length
* 76209 78970: contig of 2662 bp in length
* 78971 78970: gap of unknown length
* 81769 81769: contig of 2799 bp in length
* 81770 81869: gap of unknown length
* 81870 84942: contig of 3073 bp in length
* 84943 85042: gap of unknown length
* 85043 87086: contig of 2044 bp in length
* 87087 87186: gap of unknown length
* 87187 89756: contig of 2570 bp in length
* 89757 89856: gap of unknown length
* 89857 92029: contig of 2173 bp in length
* 92030 92129: gap of unknown length
* 92130 95030: contig of 2901 bp in length
* 95031 95130: gap of unknown length
* 95131 98584: contig of 3454 bp in length

Query Match 12.4%; Score 170.2; DB 2; Length 165337;
Best Local Similarity 50.3%; Pred. No. 5.8e-17;
Matches 418; Conservative 0; Mismatches 413; Indels 0; Gaps 0;

QY 249 CATCGTACCCAGAGAGACATCGTTACAACTCCATGTACAGTCCCTCTCCATCTC 308
Db 103750 CAACTCAACAACAGCAAGCAAAAACAGCAACAGCAACAGCAACAGCAACAA 103809

QY 309 CTTCTGGATCAGATCAACAAGTGGTCTCCAACTTGCAGGTACACCATCTCGACTC 368
Db 103810 CAACAACAGCAAGCAAAATTAACAAGCAACAAACAGCAACAACTACAACAACAGCAA 103869

QY 369 CGTCAAGAAACACTCCGGTGTGCTCCATCGTATCATCTCCAACTCTCCGTGTTCAACCT 428
Db 103870 CANTAACAACAACAGCAGCAACAACAGCAACAAACAGCAACAGCAACAGCAACAA 103929

QY 429 GAGCAGACAGGAGACTCCGACAGCTCCTCACTTCTTCTTACGACATCTCCACACAGC 488
Db 103930 CAACAACAACAGCAACAACAGCAACAGCAACAAACAGCAACAGCAACAAACAGCAACAA 103989

QY 489 TCCTGGTTACAACAAGTGGTCTTCTTGGTCCCGTCAACCGTCAACCAACACATGATGGGTAAATGAA 548

```

```

Db 103990 CAACAGCAACACACAGCAACAACAAGGCAACACAGCAACACACACACAA 104049
QY 549 GATCTACATCAACGGTAAAGCTGATCGACACCATCAAGTCAAGGAGTTGACCGGTATCAA 608
Db 104050 CAGCAACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACA 104109
QY 609 CTTCTCCAGACCATCACTCTTCAGATCAACAAGATCCAGACACCGGTCTGATCACTC 668
Db 104110 CAACAACAACAGCAACAGCAACAACAACAACAACAACAACAACAACAACAACA 104169
QY 669 CGACTCCGACAAATCAACATGTGGATCCGTGACTTCTACATCTTCGCCAAGGAGTTGGA 728
Db 104170 CAACAGCAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACA 104229
QY 729 CGGTAGGACATCAACATCTCTTCAACTCTTCGTGAGTACACCAACAGTCTCGTCAAGGACTA 788
Db 104230 CAGCAACAACAACAGCAACAGCAACAACAACAACAACAACAACAACAACAACAACA 104289
QY 789 CTGGGTAAAGACCTGAGATACAACAAGAGTACTACATGTCTACATCGACTACTTTGAA 848
Db 104290 CAGCAACAACAACAGCAACAGCAACAACAACAACAACAACAACAACAACAACAACA 104349
QY 849 CAGATACATGTAGCCCACTTCAGACAGATCGTCTTCAACACAGACGTACACACAGCA 908
Db 104350 CAACAACAACAGCAACAGCAACAACAACAACAACAACAACAACAACAACAACAACA 104409
QY 909 CTTCAACGAGGGTTACAAGATCATCATCAAGCGTATCAGAGTAAACCAACAACAGCACAG 968
Db 104410 CRACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 104469
QY 969 AGTCAGAGTGGTGGATCTCTTACTTGGACATGACTATCAACAACAAGGCTTCAACCT 1028
Db 104470 CRAACAGCAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 104529
QY 1029 GTTCTGAAGAAGAGAGACCATGTACGCGCAACCAACCACTCCACCGAGGACAT 1079
Db 104530 CRAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 104580

RESULT 15
LOCUS AR000030 1402 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 25 from patent US 5736139.
ACCESSION AR000030
VERSION AR000030.1 GI:3962561
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1402)
AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.
TITLE Treatment of Clostridium difficile induced disease
JOURNAL Patent: US 5736139-A 25 07-APR-1998;
FEATURES
    Location/Qualifiers
        source
            1..1402
            /organism="unknown"
BASE COUNT 420 a 360 c 260 g 362 t
ORIGIN

Query Match 12.4%; Score 169.4; DB 6; Length 1402;
Best Local Similarity 54.4%; Pred. No. 1.4e-16;
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;

QY 41 CCAACAACATCCCTGTTGAGGACATCATCAAGAGTACTTCAACAACATCAACGACTCCA 100
Db 68 CTAGCATGCTCGTCTGCTGTCTACTCTCACTGATACATCATCAAGACATCATCAATACCT 127

QY 101 AGATCCTGCTCCGTGAGACACCGTAAAGACACCTTGGTCGACACCTCCGGTTACAAACGCG 160
Db 128 CCATCCTGAACCTGGCTAGCAATCAATCACTGATCGACCTGTCTCGCTACGCTTCCA 187

QY 161 AGGTCTCCGAGGAGGTGAGCTCCAGCTGAACCCATCTTCCCATTCGACTTCAAGCTGG 220

```

```

Db 188 AAATCAACATCGGTCTAAAGTTAACTTCGATCGGACAGAATCAGATCCAGCTG- 246
QY 221 GTTCTCCGGTGAGGACAGAGGTAAAGTTCATCGTACCCAGAGACGAGACATCGTCTACA 280
Db 247 -----TTCAATCTGGAACTTCCAAAATCGAAGTTATCCTGAAGAATGCTATCGTATACA 301
QY 281 ACTCATGTAGAGTCCCTTCCTCCATCTCCTCTGGATCAGATCAACAACAGTGGTCTCCA 340
Db 302 ACTCTATGTAGAAAATCTCCACCTCCCTCTGGATCCGATATCCCGAATACTTCAACT 361
QY 341 ACTTGCC-----AGGTTACACCATCATCGACTCCGTCAGAAACAACATCCGGTTGGT 391
Db 362 CCATCTCTCTGAACAATGAATACACCATCATCAACTGCATGGAAAACAATCTGGTTGGA 421
QY 392 CCATCGGTATCATCTCCAACHTCTCTGGTCTTCACCTTGAGCAGACGAGGACTCCGAGC 451
Db 422 AAGTATCTCTGAACATACGCTGAAATCATCTGGACTCTGCAGGACACTCAGGAAATCAAAC 481
QY 452 AGTCCATCAACTTCTCCTACGACATCTCCAAACAGCTCCTGGTTACAACAA---GTGGT 508
Db 482 AGCGTGTGTTATTCAAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGGA 541
QY 509 TCTTCGTACCGGTACCAACAACATGATGGGTAACTGAGATCTACATCAACGGTAAAC 568
Db 542 TCTTCGTTACCATCACCAACAATGCTGTGATATACTCCAAAATCTACATCAACGGCCGTC 601
QY 569 TGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAGACCATCACCT 628
Db 602 TGATCGACGAGAACCGATCTCCATCTGGGTACATCCACGCTTCTTAATACATCATGT 661
QY 629 TCGAGATCAACAAGATCCGACAGACCGGTCTGTATCACTCCGACTCCGACACATCAACA 688
Db 662 TCAAACTGGACGTTGTGCTGACATC-----ACCGGTACA 697
QY 689 TGTGGATCGGTGACTCTTACATCTTCGCCAAGGAGTTGGAGGTAAAGACATCAACATCC 748
Db 698 TGTGGATCAAAATCTTCAATCTGTTCGACAAAGACTGACGAAAAGAAATCAAGACC 757
QY 749 TGTTCAACTCCTTGCAGTACACCAACGTCGTCAGGACTACTGGGGTAAAGACTGAGAT 808
Db 758 TGTACGACAAACCAAGTCCAATCTGTGATCTCCTGAAAGACTTCTGGGGTGAATCTGAGT 817
QY 809 ACAACAAGGAGTACTACATGTGATCGACTACTTGAACAGATACATGTACGCCAAC 867
Db 818 ACGACAAACCGTACTACATGTGAATCTGTAGATCCGACAAATACGTTGACGTCAC 876

```

Search completed: November 7, 2002, 13:24:47  
Job time : 3520 secs

GenCore version 5.1.3  
 Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 11:02:35 ; Search time 253 Seconds  
 (without alignments)  
 12203.517 Million cell updates/sec

Title: US-09-910-186A-9  
 Perfect score: 1371  
 Sequence: 1 gaattcagcatgaccatccc.....tctccgagtaataaggaaattc 1371

Scoring table: IDENTITY.NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
 Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

## Database : N\_Geneseq\_101002.\*

1: /SID82/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
 2: /SID82/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
 3: /SID82/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
 4: /SID82/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
 5: /SID82/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
 6: /SID82/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
 7: /SID82/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
 8: /SID82/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
 9: /SID82/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
 10: /SID82/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
 11: /SID82/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
 12: /SID82/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
 13: /SID82/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
 14: /SID82/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
 15: /SID82/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
 16: /SID82/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
 17: /SID82/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
 18: /SID82/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
 19: /SID82/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
 20: /SID82/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
 21: /SID82/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
 22: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
 23: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
 24: /SID82/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1371	100.0	1371	21	AA87214
2	1371	100.0	1371	22	AA87214
3	665.2	48.5	3950	20	AA87214
4	649.8	47.4	1502	19	AA87214
5	526.4	38.4	1208	21	AA87214
6	526.4	38.4	1440	21	AA87214
7	420.2	30.6	1374	22	AA87214
8	217.4	15.9	1347	21	AA87214
9	217.4	15.9	1368	22	AA87214

10	199.8	14.6	1317	22	AA87214	Botulism toxin hea
11	198.2	14.5	1317	21	AA87214	DNA encoding synth
12	169.6	12.4	1351	19	AA87214	Clostridium botuli
13	169.4	12.4	1323	22	AA87214	Botulism toxin hea
14	169.4	12.4	1402	19	AA87214	Type A neurotoxin
15	169.4	12.4	1402	17	AA87214	Clostridium botuli
16	168.8	12.3	1326	22	AA87214	Botulism toxin hea
17	168.8	12.3	1330	17	AA87214	Type A neurotoxin
18	168.8	12.3	1330	19	AA87214	Clostridium botuli
19	168.8	12.3	1332	22	AA87214	Botulism toxin hea
20	168.8	12.3	1338	21	AA87214	DNA encoding synth
21	162.4	11.8	1400	21	AA87214	Botulism toxin hea
22	162.4	11.8	1400	22	AA87214	DNA encoding synth
23	157.4	11.5	1278	22	AA87214	Botulism toxin hea
24	151.8	11.1	1314	22	AA87214	Botulism toxin hea
25	151.6	11.1	1313	18	AA87214	Immunogenic type F
26	147.2	10.7	985	24	ABQ44816	Oligonucleotide fo
27	147.2	10.7	985	24	ABQ44816	Oligonucleotide fo
28	145.8	10.6	1332	22	AA87214	Sequence encoding
29	136	9.9	5163	19	AA87214	Cryptosporidium pa
30	136	9.9	5163	21	AA87214	ORF encoding a por
31	136	9.9	5163	21	AA87214	C parvum Gp900 gen
32	136	9.9	5318	24	AA87214	DNA encoding a por
33	136	9.9	5318	24	AA87214	C parvum Gp900 gen
34	134.4	9.8	5318	19	AA87214	Cryptosporidium pa
35	133.6	9.7	7417	23	ABL16368	Drosophila melanog
36	126.6	9.2	702	21	AA87214	DNA encoding BONTA
37	122.8	8.0	1347	22	AA87214	Sequence encoding
38	121	8.8	2607	23	ABL16368	Drosophila melanog
39	120.4	8.8	1341	21	AA87214	DNA encoding synth
40	120.4	8.8	1341	22	AA87214	Botulism toxin hea
41	115.4	8.4	5511	21	AA87214	Cryptosporidium pa
42	115.4	8.4	5511	24	AA87214	C parvum Gp900 gen
43	115.4	8.4	7334	24	AA87214	Cryptosporidium pa
44	115.4	8.4	7334	24	AA87214	C parvum Gp900 gen
45	110.6	8.1	3297	23	AA87214	DNA encoding novel

## ALIGNMENTS

RESULT 1  
 ID AA87214 standard; DNA; 1371 BP.  
 AC AA87214;  
 XX  
 DT 08-MAY-2000 (first entry)  
 DT  
 DE DNA encoding synthetic BONT serotype C (BONT) HC fragment.  
 KW Botulinum neurotoxin; heavy chain; BONT; serotype C;  
 KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 KW VEE; botulism; vaccine; diagnosis; drug screening; ds.  
 XX  
 OS Clostridium botulinum.  
 OS Synthetic.  
 XX  
 FH Key  
 FT CDS  
 FT  
 FT Location/Qualifiers  
 FT 10..1362  
 FT /\*tag= a  
 FT /product= "Synthetic botulinum neurotoxin serotype C  
 (BONT) heavy chain C-terminal fragment (HC)"  
 XX  
 PN WO200002524-A2.  
 XX  
 PD 20-JAN-2000.  
 XX  
 PF 09-JUL-1999; 99WO-US15570.  
 XX  
 PR 10-JUL-1998; 98US-0092416.  
 PR 12-MAY-1999; 99US-0133870.  
 XX





```

KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
FH Key Location/Qualifiers
CDS CDS 10..1362
FT /*tag= a
FT /product= H_C peptide fragment
XX
PN WO200067700-A2.
XX
XX 16-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US12890.
XX
XX 12-MAY-1999; 99US-0133865.
XX
XX 12-MAY-1999; 99US-0133866.
XX
XX 12-MAY-1999; 99US-0133867.
XX
XX 12-MAY-1999; 99US-0133868.
XX
XX 12-MAY-1999; 99US-0133869.
XX
XX 12-MAY-1999; 99US-0133873.
XX
XX 29-JUL-1999; 99US-0146192.
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
XX WPI; 2001-016048/02.
XX
XX P-PSDB; AAB04092.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as
XX vaccine against botulinum
XX
XX Claim 2; Fig 5a; 73pp; English.
XX
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide
XX chain and then posttranslationally nicked, forming a dichain
XX consisting of a 100 kDa heavy chain and a 50 kDa light chain which
XX remain linked by a disulfide bond. Nucleic acids encoding the
XX carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
XX chain of botulinum neurotoxin (BoNT) can be used in recombinant
XX expression vectors and expressed in transfected cells to produce
XX peptide antigens useful for eliciting an immune response to give
XX protective immunity against botulinum neurotoxin, which causes
XX botulism. The nucleic acids are expressible in a recombinant
XX organisms such as Escherichia coli or Pichia pastoris. The use
XX of recombinant nucleic acids are advantageous since it eliminates
XX the need to culture large quantities of hazardous toxin-producing
XX bacterium. Production yield from the genetically engineered product
XX is also high and cost of production is lower. The nucleic acids can
XX be derived from Clostridium botulinum serotypes A-G.
XX
XX Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;
XX
XX Query Match 100.0%; Score 1371; DB 22; Length 1371;
XX Best Local Similarity 100.0%; Pred. NO. 4,1e-289;
XX Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps
XX 0;
XX
XX 1 GAATTACGATGACATCCCATCCATCTCTCTACACCAACAACTCCCTGTTGAAG 60
XX |
XX 1 GAATTACGATGACATCCCATCCATCTCTCTACACCAACAACTCCCTGTTGAAG 60
XX
XX 61 GACATCATCAACGAGTACTTCAACACATCAACGACTCCCAAGATCTGTCCTGCAGAAC 120
XX |
XX 61 GACATCATCAACGAGTACTTCAACACATCAACGACTCCCAAGATCTGTCCTGCAGAAC 120
XX
XX 121 CGTAGAACACTTGGTGGACACTCGGTTACACGGCGAGGTCTCCGAGGAGGTGAC 180
XX |
XX 121 CGTAGAACACTTGGTGGACACTCGGTTACACGGCGAGGTCTCCGAGGAGGTGAC 180
XX
XX 181 GTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCTCTCCGGTGAGGACAGA 240
XX |

```

```

181 GTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCTCTCCGGTGAGGACAGA 240
241 GGTAAAGTCTATCGTACCCAGAGACGAGACATCGTGTACAACTCCATGTAGAGTCCCTTC 300
241 GGTAAAGTCTATCGTACCCAGAGACGAGACATCGTGTACAACTCCATGTAGAGTCCCTTC 300
301 TCCATCTCCTCTCTGGATCAGAATCAACAAGTGGGTCTCCAACTTGGCCAGGTACACATC 360
301 TCCATCTCCTCTCTGGATCAGAATCAACAAGTGGGTCTCCAACTTGGCCAGGTACACATC 360
361 ATCGACTCCGTCAAGAACAACTCCGGTGGTTCATCGGTATCATCTTCCAACTTCCGTGTC 420
361 ATCGACTCCGTCAAGAACAACTCCGGTGGTTCATCGGTATCATCTTCCAACTTCCGTGTC 420
421 TTCACCTCTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCC 480
421 TTCACCTCTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCC 480
481 AACACGCTCTCTGGTTACAAACAAGTGGTTCCTTCCGTCCCGTCAACCAACATGATGGT 540
481 AACACGCTCTCTGGTTACAAACAAGTGGTTCCTTCCGTCCCGTCAACCAACATGATGGT 540
541 AACATGAGATCTACATCAACGATGAGTGCAGACCATCAAGTCAAGGTCAAGGATGGAC 600
541 AACATGAGATCTACATCAACGATGAGTGCAGACCATCAAGTCAAGGTCAAGGATGGAC 600
601 GGTATCAACTTCTCCAGACCATCACTTTCGAGATCAACAAGATCCCGACACCCGCTG 660
601 GGTATCAACTTCTCCAGACCATCACTTTCGAGATCAACAAGATCCCGACACCCGCTG 660
661 ATCACCCTCCGACTCCGACAACTCAACATGTGGATCCGTTGACTTCTACATCTTCCGCAAG 720
661 ATCACCCTCCGACTCCGACAACTCAACATGTGGATCCGTTGACTTCTACATCTTCCGCAAG 720
721 GAGTTGACGCTAAGGACATCAACATCTCTTCAACTCTCTGAGTACACCAACGCTGTC 780
721 GAGTTGACGCTAAGGACATCAACATCTCTTCAACTCTCTGAGTACACCAACGCTGTC 780
781 AAGSACTACTTGGGTTACGACCTGAGATACACAAAGAGTACTACATGTCTCAACATCGAC 840
781 AAGSACTACTTGGGTTACGACCTGAGATACACAAAGAGTACTACATGTCTCAACATCGAC 840
841 TACTTGAACAGATACATGTAGCGCACTCCAGACAGATCGTCTTCAACACCAAGAGTAAAC 900
841 TACTTGAACAGATACATGTAGCGCACTCCAGACAGATCGTCTTCAACACCAAGAGTAAAC 900
901 AACACGACTTCAACGAGGTTACAGATCATCATCAAGGTTATCAGAGTTACACCAAC 960
901 AACACGACTTCAACGAGGTTACAGATCATCATCAAGGTTATCAGAGTTACACCAAC 960
961 GACACCAAGTTCAGAGGTTGAGTCTCTGACTTCCGATGACTATCAACCAACAGGCC 1020
961 GACACCAAGTTCAGAGGTTGAGTCTCTGACTTCCGATGACTATCAACCAACAGGCC 1020
1021 TACAACCTGTTTATGAGACGAGACCATGTAGCGCGCAACCACTCCACCGAGGACATC 1080
1021 TACAACCTGTTTATGAGACGAGACCATGTAGCGCGCAACCACTCCACCGAGGACATC 1080
1081 TACGCCATCGGTCTGCGTGAGCAGACCCAGGACATCAACGACACATCATCTTCCAGATC 1140
1081 TACGCCATCGGTCTGCGTGAGCAGACCCAGGACATCAACGACACATCATCTTCCAGATC 1140
1141 CAGCCAATGAACAACACTTACTACTACGCTTCCAGATCTTCAAGTCCCACTTCAACGGT 1200
1141 CAGCCAATGAACAACACTTACTACTACGCTTCCAGATCTTCAAGTCCCACTTCAACGGT 1200
1201 GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATTCCTGCTGGTGGTGGTAC 1260
1201 GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATTCCTGCTGGTGGTGGTAC 1260
1261 TACAGACAACTACTTGGTTCCTCAACTGTCAAGCAGGGAATACGCTTCCTTGGTGGAG 1320
1261 TACAGACAACTACTTGGTTCCTCAACTGTCAAGCAGGGAATACGCTTCCTTGGTGGAG 1320

```

Db 1261 TACAGACACAACTACTTGGTTCACACTGTCAACAGAGGTAACTACGCCCTCCCTGGTGGAG 1320  
 QY 1321 TCCACTCCACCCACTGGGATTCGTCGCCAGTCTCCGAGTAAATAGAAATTC 1371  
 Db 1321 TCCACTCCACCCACTGGGATTCGTCGCCAGTCTCCGAGTAAATAGAAATTC 1371

RESULT 3  
 AAX25521  
 ID AAX25521 standard; DNA; 3950 BP.  
 AC AAX25521;  
 XX  
 XX 02-AUG-1999 (first entry)  
 DT DNA coding for modified botulinum toxin rBont/C.  
 DE Botulinum toxin; botulism; rBont/C; vaccine; drug delivery;  
 XX mutant; ss.  
 KW Clostridium botulinum.  
 XX Synthetic.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 32..3907  
 FT /\*tag= a  
 PN W09920306-A1.  
 XX  
 XX 29-APR-1999.  
 PD  
 XX  
 XX 16-OCT-1998; 98WO-US21897.  
 PF  
 XX  
 XX 20-OCT-1997; 97US-0954302.  
 PR  
 XX  
 XX (UYJE-) UNIV JEFFERSON THOMAS.  
 PA  
 XX  
 XX Kiyatkin N, Maksymowych A, Simpson L;  
 PI  
 XX  
 XX WPI; 1999-302646/25.  
 DR  
 XX P-PSDB; AAY05814.  
 XX  
 XX Modified toxin useful for systemic delivery of oral vaccines and  
 PT therapeutic agents  
 PT  
 XX  
 XX Example 1; Page 26-31; 37pp; English.  
 PS  
 XX  
 XX This DNA sequence codes for a modified serotype C botulinum toxin,  
 CC termed rBont/C (see AAY05814), in which amino acids His-229, Glu-230  
 CC and His-233 of the native sequence are substituted by Gly, Thr and  
 CC Asn, respectively, i.e. the zinc binding motif (see AAY05817) of  
 CC the light chain holotoxin is modified, resulting in loss of  
 CC endoprotease activity. DNA coding for the modified botulinum toxin  
 CC was assembled from 3 separate toxin fragments using PCR and  
 CC site-directed mutagenesis. The modified recombinant botulinum  
 CC toxin maintains its ability to translocate from the gut into the  
 CC general circulation but is non-toxic. It can be used as an oral  
 CC vaccine for antigenic peptides, including botulinum toxin (i.e. an  
 CC oral vaccine for botulism) or for the oral delivery of other  
 CC therapeutic agents to the general circulation.  
 XX  
 XX Sequence 3950 BP; 1589 A; 410 C; 616 G; 1335 T; 0 other;  
 SQ

Query Match 48.5%; Score 665.2; DB 20; Length 3950;  
 Best Local Similarity 68.3%; Pred. No. 3e-135;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTCTCTACACCAACTCCCTGTTGGAAGACATCATCAAC 72  
 Db 2558 ACATACCCCTTTAATATTTTTCATATACTAATAATCTTTTATAAAGATATAATTAAT 2617  
 QY 73 GAGTACTTCAACACATCAACGACTCCAGACTCCCTGCTCCACGAACCGTAAGACACC 132

Db 2618 GAATATTTCAATTAATTAATGATTCAAAAAATTTTGAGCCTACAAAAACAGAAAAAATACT 2677  
 QY 133 TTGCTCGACACCTCCCGGTTAAACGGCGAGTCTCCGAGAGGGTGACCTCCAGCTGAC 192  
 Db 2678 TTAGTGGATACATCAGGATATAATGCAAGAGTGAAGAAGGGGATGTTCCAGCTTAAT 2737  
 QY 193 CCAATCTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGAGTAAAGTCAATC 252  
 Db 2738 CCAATATTTCCATTGACTTAAATAGGTAGTTTCAGGGGAGGATAGAGTAAAGTTATA 2797  
 QY 253 GTCACCCAGAACGAGAACATCGTCTCAACATCCATGATAGAGTCTCTCCATCTCCCTTC 312  
 Db 2798 GTAACCCAGAAATGAAAAATATTGTATATAATTTCTATGATCAAAAGTTTATAGCATTT 2857  
 QY 313 TGGATCAGAACTCAACAAAGTGGTCTCCAACTTCGCCAGGTTCACCACTATCGCACTCCGTC 372  
 Db 2858 TGGATTAGATAATAATAATGGTAAGTAATTTACCTGGATATATCTATTAATTGATAGTGT 2917  
 QY 373 AAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCTGGTCTTTCACCCCTGAG 432  
 Db 2918 AAAAATAACTCAGGTGGAGTATAGGTATTATTAGTAATTTTTTTAGTATTACTTTTAAAA 2977  
 QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCCACAAACGCTCCT 492  
 Db 2978 CAAAATGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAATAATGCTCCT 3037  
 QY 493 GGTACAAACAAAGTGTCTTCCTCCGTCACCGTCAACCAACAATGATGGGTACATGAAGATC 552  
 Db 3038 GGATACAATAAATGGTTTTTTTGAATTTACTAACAAATATGATGGGAAATATGAAGATT 3097  
 QY 553 TACATCAACGGTGAAGTCTCGACACCATCAAGTCAAGGAGTTGACCGGTATCAACTTC 612  
 Db 3098 TATATAAATGGAAAAATTAATAGATACTATAAAGTTAAAGTAACACTGGAATTAATTTT 3157  
 QY 613 TCCAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCTGTATCACTCCGAC 672  
 Db 3158 AGCAAACTATAACATTTGAATAAATAAATTCAGATACCGGTTTCATTACTTCAGAT 3217  
 QY 673 TCCGACACATCAACATGTGGATCCGTGACTTCTACATCTTCGCCAGAGGATTTGGAGGT 732  
 Db 3218 TCTGATACATCAATATGTGGATAGAGATTTTTTATATATTTGCTAAAGAAATAGATGGT 3277  
 QY 733 AAGGACATCAACATCTCTTCAACTCTTCAGTACACCAACGCTGCTCAAGGACTACTGG 792  
 Db 3278 AAAGATATATATATTTTAAATAGCTTGCAATATACTAATGTTGTAAAGATTATTGG 3337  
 QY 793 GGTAACGACCTGAGATACAACAAGGAGTACTACATGGTCACTACCTGACTTCAACAGA 852  
 Db 3338 GGAAATGATTTAAGATATAATAAAGAAATATTATGTTAATATAGATTATTAAATAGA 3397  
 QY 853 TACATGTAGCCCACTCCAGACAGATCTCTTCAACACCAAGACGCTAAACAACGACTTC 912  
 Db 3398 TATATGTATGCGAATCAGACAAATTTCTTTTAACACGTAGAAATAATAATGACTTC 3457  
 QY 913 AACGAGGTTCAAGATCATCATCAAGCGTATCAGAGTACACCAACGACCAAGATC 972  
 Db 3458 AATGAAGGATATAAAATTTATAATAAAGAAATCAGAGAAATCAAAATGATACTAGAGTA 3517  
 QY 973 AGAGGTGGTACATCTCTGACTTCGACATGACTATCAACAACAAGGCTTCAACCTGTC 1032  
 Db 3518 CGAGGAGGAGATTTTTATATTTTGTATGACAAATTAATAACAAGCATATAATTTGTTT 3577  
 QY 1033 ATGAAGAAGACGACCATGTACGCCCAACCAACCTCCACCGAGGACATCTAGCCCATGGT 1092  
 Db 3578 ATGAAGAATGAACATATGTATGATGATATATCATAGTACTGAAGATATATATGCTATAGT 3637  
 QY 1093 CTGCGTGACGACCAAGGACATCAACGACACATCACTTCCAGATCCAGCCCAATGAAC 1152  
 Db 3638 TTAAGAGAACAAACAAAGGATATAAATGATATATATATTTCAATACCAACCAATGAAT 3697  
 QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAAGGGTGAGAACATCTCC 1212  
 Db 3698 AATACATTTATTAACGATCTCAAAATTTTAAATCAAAATTTTAAATGGAGAAAAATTTTCT 3757



Db 1195 ATCATAGTACTGAGATATATATGCTATAGGTTTAAAGACAAACAAGGATATAAATG 1254  
 Qy 1121 ACAACATCATCTCCAGATCCAGCCAAATGACAAACACTTACTACTAGCTTCCCAAGATCT 1180  
 Db 1255 ATAATATATATTTCAATACAAACCAATGAATAATATTTATTACGCATCTCAAAAT 1314  
 Qy 1181 TCAAGTCCCACTTCAACGGTGAGAACATCTCCGGTATCTGTTCCATCGGTACTACAGAT 1240  
 Db 1315 TTAATCAATTTTAAATGGAGAAATATTTCTGGAAATATGTTCAATAGTACTTATCGTT 1374  
 Qy 1241 TCCGTCGTGGGTGAGTCTGATCAGACAACTACTTGTGTTCCAACTGTCAAGCAGGGA 1300  
 Db 1375 TTAGACTTGGAGGTGATTGGTATAGACAAATATTTTGGTGGCTACTGTGAAGCAAGAA 1434  
 Qy 1301 ACTAGCCTCTCTGCTGAGTCCAGTCCACCCACTGGGATTCGTCCTCAGTCTCCGAGT 1360  
 Db 1435 ATTATGCTCTATTTAGATCAACATCACTCATTTGGGGTTTGTACCTGTAAAGTGAAT 1494  
 Qy 1361 AATAG 1365  
 Db 1495 AAAAG 1499

RESULT 5  
 AAZ98630  
 ID AAZ98630 standard; DNA; 1208 BP.  
 XX  
 AC  
 XX  
 XX  
 DT  
 XX  
 XX  
 DE C. botulinum type D toxin nucleotide sequence.  
 XX  
 KW Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;  
 KW protect; ds.  
 XX  
 OS Clostridium botulinum.  
 XX  
 PN WO200005252-A1.  
 XX  
 PD 03-FEB-2000.  
 XX  
 XX 20-JUL-1999; 99WO-IB01301.  
 PF  
 XX 22-JUL-1998; 98ZA-0006538.  
 PR  
 XX (AGRI-) AGRIC RES COUNCIL.  
 PA  
 XX De Bruyn EE, Botha AD;  
 PI  
 XX WPI: 2000-205375/18.  
 DR P-PSDB; AAY78982.  
 XX  
 XX Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin  
 PT type D, useful in vaccines for protection against botulism, comprises  
 PT at least one amino acid mutation not present in the wild type D  
 PT neurotoxins  
 XX  
 PS Claim 8; Page 54-57; 66pp; English.

XX This sequence represents a nucleotide sequence of a synthetic gene  
 CC encoding a non-toxic immunogenic derivative of Clostridium botulinum type  
 CC D toxin (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle  
 CC and sheep, and usually results in the death of the affected or poisoned  
 CC animal. The non-toxic immunogenic fragments of the C. botulinum  
 CC neurotoxin are useful in vaccines to protect animals (e.g. humans,  
 CC cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic  
 CC fragments can be produced relatively simply and inexpensively  
 CC (specifically by fermentation techniques). As the fragments are not  
 CC toxic the risk to production staff is reduced.  
 XX  
 XX Sequence 1208 BP; 416 A; 203 C; 216 G; 373 T; 0 other;

Query Match 38.4%; Score 526.4; DB 21; Length 1208;  
 Best Local Similarity 65.9%; Pred. No. 3.7e-105;  
 Matches 799; Conservative 0; Mismatches 401; Indels 12; Gaps 2;  
 Qy 157 GCGAGGTCTCCGAGGAGGTGACGTCCAGCTGAACCCAACTTCCCATTCGACTTCAAG 216  
 Db 4 GCAGAAATAGAGTTTGAAGGAAACGTGAGTGAATCCCTATCTTCCCATTTGACTTAAAG 63  
 Qy 217 CTGGTTCCTCCCGTGGAGGACAGAGTAAGGTTCATCGTCACOCAGAGAGAGAACATCGTC 276  
 Db 64 TTGGGTAGCTCTGCTGATGACCGTGCAGAAATATTTGTAACCTCAAAACGAAATATTTGT 123  
 Qy 277 TACAACCTCCATGACAGTCTCTCCATCTCCTCTGATCAGATCAACAGTGGGTC 336  
 Db 124 TATAATGCTATGATGAGTCTTCTATAGCTTCTGATGATGATTAATAATGAGTGGT 183  
 Qy 337 TCAACCTGCGAGGTGACCATCATCGATCCGTCAGAGCAAACTCCGGTTGGTCCATC 396  
 Db 184 TCAATTTACAGGTATACCATCATCGACGGTTAAATAATTCAGGATGGTCTATT 243  
 Qy 397 GGTATCATCTCCAACTTCTCTGCTTCCCTGACGACAGACGAGGACTCCGAGCAGTCC 456  
 Db 244 GGCATAATTAGCAATTTTCTAGTTTACACTAAAGCAAAAGAAATAGCGAGCAGGAC 303  
 Qy 457 ATCAACTTCTCTACGACATCTCCAAACCGCTCTCTGGTTTACAAAGAGTGGTCTTCGTC 516  
 Db 304 ATTAATTTTCTTATGACATTTCAAGAGCGCTGCTGGTTATAATAATGTTCTTTGTA 363  
 Qy 517 ACCGTCAACCAACATGATGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 576  
 Db 364 ACCATTACCAACATGATGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 423  
 Qy 577 ACCATCAAGGTCAAGGAGTGGACGGGTATCACTTCTCCAAAGACCATCACTTCGGAGATC 636  
 Db 424 ACCATCAAGGTCAAGGAGTGGACGGGTATCACTTCTCCAAAGACCATCACTTCGGAGATG 483  
 Qy 637 AACAGATPCCAGACACCGGTCTGATCACCTCCGATCCGACACATCAACATGTCGATC 696  
 Db 484 AATAAAATCCCGAACACAGGATTAATAACGTCTGATCTGACAACTCAATATGTTGAT 543  
 Qy 697 CGTGACTTCTACATCTTCCGCAAGGAGTGGACGGTAAAGGACATCAACATCCCTGTTCAAC 756  
 Db 544 AGAGATTTCTATATTTTCCGCAAGGAGTGGACGGTAAAGGACATCAACATCCCTGTTCAAC 603  
 Qy 757 TCCCTTGAGTACACCAACGTGCTCAAGGAGTCTGAGGTAAACACCTGAGATCAACACAG 816  
 Db 604 TCACCTCAGTACACTAATGTTGTTAAGGATTAATGGGTAATGATTTAAGATATGATAAG 663  
 Qy 817 GAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGTAGCGCAACTCCAGACAG 876  
 Db 664 GAGTACTACATGATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 723  
 Qy 877 ATCGTCTTCAACACAGACGTAAACAAACGACTTCAACGAGGTAAAGAGTCAACATCATC 936  
 Db 724 ATCGTCTTCAACACGTAAAGTAAACAAACGATTTCAACGAGGTAAAGAGTCAACATCATC 783  
 Qy 937 AAGCGTATCAGAGGTAAACAAACGACACGAGTCAAGGTGAGTGTGATCCTGTTACTTC 996  
 Db 784 AAAGAAATCCCGTGAATACAAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 843  
 Qy 997 GACATGATATCAACAAACGAGGCTTCAACCTGTTCTATGAAGACGAGACCATGTAGGCC 1056  
 Db 844 AATAACAGATAGATAATAAGCAGTACTCTTTAGGCATGTATAAAGCGTCAA----- 895  
 Qy 1057 GACACCACTCCACGAGGACATCTACCCCTGCTGCGTGGAGCAGACCAAGGACATC 1116  
 Db 896 -GAAACTTAGGACAGACCTGTTTCCATGGGTGATATAGATCAACGATGGAGCAAT 954  
 Qy 1117 AACGCAACATCATCTTCCAGATCCAGCAATGAACACACTTACTACTAGGTTCCAG 1176  
 Db 955 CGTAATATGGATCATTCATTAATTAACCATGTAATAGCTTCGATTAATAGCGAGCAA 1014  
 Qy 1177 ATCTTCAAGTCAACTTCAACGGTGAAGACATCTCCGGTATCTCTTCCATCGGTACCTAC 1236



QY 1294 CAGGCTAACTAGCGCTCTCTGCTGAGTCCACTCCACCCAGCTGGGATTCGTCACAGTC 1353  
 Db 1189 ATCGAATACGCTTCAATTAAGTCACTCTTACACATTTGGGTCTTGTCCAGCG 1248  
 QY 1354 TCCGAGTAATAG 1365  
 Db 1249 TCTGAGTAAAG 1260

RESULT 7  
 AAA54487  
 ID AAA54487 standard; DNA; 1374 BP.  
 XX  
 AC AAA54487;  
 XX  
 DT 11-APR-2001 (first entry)  
 XX Botulinum toxin heavy chain C-terminal coding sequence (serotype D).

XX Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;  
 KW recombinant vector; antigen; immune response; vaccine; bacterium;  
 KW infection; ds.

XX Synthetic.  
 OS Clostridium botulinum.  
 XX  
 PH Key Location/Qualifiers  
 CDS 10..1365  
 FT /\*tag= a  
 FT /product= H\_C peptide fragment  
 FT  
 PN WO200067700-A2.

XX 16-NOV-2000.  
 XX  
 PF 12-MAY-2000; 2000WO-US12890.  
 PR 12-MAY-1999; 99US-0133865.  
 PR 12-MAY-1999; 99US-0133866.  
 PR 12-MAY-1999; 99US-0133867.  
 PR 12-MAY-1999; 99US-0133868.  
 PR 12-MAY-1999; 99US-0133869.  
 PR 12-MAY-1999; 99US-0133873.  
 PR 29-JUL-1999; 99US-0146192.  
 XX

PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX WPI; 2001-016048/02.  
 DR P-PSDB; AAB04093.

XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 PT vaccine against botulism

XX Claim 2; Fig 6a; 73pp; English.

XX Botulinism neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dichain  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulism. The nucleic acids are expressible in a recombinant  
 CC organism such as *Escherichia coli* or *Pichia pastoris*. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product

CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.

XX Sequence 1374 BP; 367 A; 423 C; 274 G; 310 T; 0 other;

QY Query Match 30.6%; Score 420.2; DB 22; Length 1374;  
 Best Local Similarity 63.0%; Pred. NO. 5.1e-82;  
 Matches 754; Conservative 0; Mismatches 398; Indels 45; Gaps 5;

QY 4 TTCAGGATGACCATCCCATTCACATCTCTCTACACCAACCACTCCCTGTTGAAGGAC 63  
 Db 40 TTCGAGAACACCATGCCATTCACATCTCTCTACACCAACCACTCCCTGTTGAAGGAC 99  
 QY 64 ATCATCAACGAGTACTTCAACACATCAACGACTCTCAAGATCTGTCCTCGCAGAACCGT 123  
 Db 100 ATCATCAACGAGTACTTCAACATCCATCAACGACTCTCAAGATCTGTCCTCGCAGAACG 159  
 QY 124 AAGAACCTTGGTCGACACTCCGGTTCAACGCGAGGCTCTCCGAGGAGGTGACGTC 183  
 Db 160 AAGAACGCTTGGTCGACACTCCGGTTCAACGCGAGGCTCTCCGAGGAGGTGACGTC 219  
 QY 184 CAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGAGT 243  
 Db 220 CAGTTGAACACCATCTACACCAAGGACTTCAAGTGTGCTCTCCGGTGA-----C 270  
 QY 244 AAGTCATCGTCACCGAGAGGAGACATCGTCTACAACTCCATGTACGAGTCTCTCTCC 303  
 Db 271 AAGTCATCGTCAACTTGAACACACATCTTGTATCCGCGCATCTACGAGAACTCTCT 330  
 QY 304 ATCTCCTTCTGGATCAGAAATCAACAGTGGGTCTCCAACTTGC---CAGGTTACACCATC 360  
 Db 331 GTCCTCTCTGGATCAAGATCTCCAAAGGACTTGACCAACTCCCAACAGGAGTACACCATC 390  
 QY 361 ATCGACTCCGTCAAGACACTCCGGTGGTTCATCGGTATCGTATCTCTCCACTCTCTGGTC 420  
 Db 391 ATCAACTCCATCGAGCAGAACTCCGGTGGAACTGTGTATCCGTAACGGTAACATCGAG 450  
 QY 421 TTCACCTTGAAGCAGAGGAGGACTCCGAGCAGTCCATCAACTTCTCTCCACGACATCTCC 480  
 Db 451 TGGATCTTCAGGAGGCTCAACCGTAAGTCAAGTCTCTGTATCTTCGACTACTCCGAGTCC 510  
 QY 481 AACAAAGCTCTCTGGTTA---CAACAAGTGGTTCCTGTCACCGTCCACCAACACATGATG 537  
 Db 511 TTGTCCACACACCGGTTACACCAACAGTGGTTCCTGTCACCGTCCACCAACACATGATG 570  
 QY 538 GGTACATGAAGATCTACATCAACGGTAAAGTGTATCGACACCATCAAGGTCAAGGAGGTG 597  
 Db 571 GGTACATGAAGTGTATCATCAACGGTAAAGTGTATCGACACCATCAAGGTCAAGGAGGTG 630  
 QY 598 ACCGGTATCAACTCTCCAGACCATCACTTCGAGATCAACAGATCCCGAGAGTCCGAGCCGT 657  
 Db 631 GACGAGGTCAAGCTGGACAGACCATCGTCTTCGGTATCGACGAGGATCGAGGAGTCCGA 683  
 QY 658 CTGATCACTCCGACTCCGACACACATCAACATCGGTATCGGTACTCTTACATCTCTCGCC 717  
 Db 684 -----CGAGAACCATGTTGGGATCGGTACTCTTACATCTCTCTCTCC 726  
 QY 718 AAGGAGTTGGAGCTAAGGACATCAACATCTCTGTTCAACTCTCTGCGAGTACACCAAGTC 777  
 Db 727 AAGGAGTGTCCAACAGGAGGACATCAACATCTCTGACGAGGTCAGATCTCTGAGGAGGTC 786  
 QY 778 GTCAGGAGTACTGGGTAAAGCACTGAGATACAAAGGAGTACTACATGTTCAACATC 837  
 Db 787 ATCAAGGAGTACTGGGTAAAGCACTGAGTTCGACCCGAGTACTACATCAACGAGC 846  
 QY 838 GACTACTTGAACAGATACATGTAGCCAACTCCAGACAGATCTCTTCAACACGAGCGT 897  
 Db 847 AACTACATCGACCGTATACATCGCCAGAGTCCCAAGCTCTCTGCTGGTCCAGTACCCT 906  
 QY 898 AACAAACAGCTTCAACGAGGTTCAAGATCATCATCAAGCGTATCAAGGATACAC 957  
 Db 907 CACCGTTCCAAAGCTGTACACCGGTAACCGCTATCACCATTCAAGTCCCTCTCCGCAAGAAC 966

Qy 958 AACGACACGAGCTCAGAGGTGGTGCATCCCTCTACTTCGACATGACTATCAACACAAAG 1017  
 Db 967 CTTTACTCCCGTATCTGACAGGTGACACATCACTCTCCATCATGCTGTACACATCCCGT 1026  
 Qy 1018 GCTTACAACTGTTGATGAGAACGAGACCATGTAGCGCGACAA-----CCACTCCACC 1071  
 Db 1027 AAGTACATGATCATCCGTGACACCGACCATCTACGCCACCCAGGGTGGTACGTGTC 1086  
 Qy 1072 GAGGACATCTACCCCATCGTCTGCTGAGCAGACCAAGGACATCAACGACACATCATC 1131  
 Db 1087 CAGAACGTGTCTACGCCCTCGAAGCTGCAGCTCCACCTGGGTAAGTACGGGTATCGGTA 1146  
 Qy 1132 TTCAGATCCAGCAATGAACACACTTACTACTAGCTTCCTCCAGATCTTCAAGTCC 1188  
 Db 1147 TTCTCCATCAAGACATCGTCTCCAGAACAACTAGTCTGCCAGATCTTCTCTCC 1203  
 RESULT 8  
 ID AA287217 standard; DNA; 1347 BP.  
 AC AA287217;  
 XX 08-MAY-2000 (first entry)  
 DT  
 DE DNA encoding synthetic BoNT serotype G (BoNTG) Hc fragment.  
 XX Botulinum neurotoxin; heavy chain; BoNT; serotype G;  
 KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 KW VEE; botulinum; vaccine; diagnosis; drug screening; ds.  
 XX Clostridium botulinum.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1347  
 FT /tag= a  
 FT /product= "Synthetic botulinum neurotoxin serotype G  
 FT (BoNTG); heavy chain C-terminal fragment (Hc)"  
 FT /note= "No stop codon given in the specification"  
 XX  
 PN WO200002524-A2.  
 XX  
 PD 20-JAN-2000.  
 XX  
 PF 09-JUL-1999; 99WO-US15570.  
 XX  
 PR 10-JUL-1998; 98US-0092416.  
 PR 12-MAY-1999; 99US-0133870.  
 XX  
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 XX  
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
 XX  
 XX WPI; 2000-160827/14.  
 DR P-PSDB; AAY77139.  
 XX  
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 PT toxin serotypes A-G, is used for inducing an immune response against  
 PT botulinum -  
 XX  
 PS Disclosure; Page 47-48; 54pp; English.  
 XX  
 CC The invention relates to novel vaccines that induce a protective immune  
 CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
 CC DNA construct comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 CC serotypes A-G. In preferred embodiments of the invention, the vector is  
 CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
 CC this vector results in the production of large amounts of a protein  
 CC encoded by a sequence cloned into the replicon. The constructs are used  
 CC to produce vaccines against botulinum. The proteins can also be used as

CC diagnostic tools for the diagnosis of botulism. The transformed host  
 CC cells can be used to analyse the effectiveness of drugs and agents which  
 CC inhibit toxin effects. The vaccine currently used against botulism is  
 CC dangerous and expensive to produce, and contains formalin, which is very  
 CC painful for the recipient. Also, the vaccine is incomplete, in that only  
 CC 5 of the 7 serotypes are represented in the formulation. The novel  
 CC vaccine of overcomes these problems, as it is easily purified, and  
 CC available in large quantities. It is also expressed in the lymph nodes  
 CC for a better immune response. Sequences AA287212-287217 represent  
 CC synthetic DNA sequences encoding BoNT Hc fragments used in the present  
 CC invention. These were optimised for codon usage for expression in yeast.  
 XX  
 XX Sequence 1347 BP; 358 A; 442 C; 263 G; 284 T; 0 other;  
 Query Match 15.9%; Score 217.4; DB 21; Length 1347;  
 Best Local Similarity 52.5%; Pred No. 7.6e-38;  
 Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;  
 Qy 43 AACAATCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAACATCAAGACTTCCAA 102  
 Db 4 AAGGACACCATCTGATCCAGGTCTTCAACAACATCATCTCCACATCTCTCCAAACGCC 63  
 Qy 103 ATCTGTCTCTGACAGAACCGTAAGACACCTTGGTGCACACCTCCGCTTACAAACGCCGAG 162  
 Db 64 ATCTGTCTCTGACAGAACCGTAAGACACCTTGGTGCACACCTCCGCTTACAAACGCCGAG 123  
 Qy 163 GTCTCCGAGGAGGTGACGTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGTGGGT 222  
 Db 124 ATGAACGTCTGTCGACGATCTTCAACGACATCGGTCAACGGTCAGTTCGAAGTGAAC 183  
 Qy 223 TCTCCGGTGAGGAGGTGACGTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGTGGGT 282  
 Db 184 AACTCC-----GAGAACTCCAAACATCACGCCGCCACGATCCCAAGTTCGTCTACGAC 237  
 Qy 283 TCCATGTACGAGTCTTCTCCATCTCTTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 342  
 Db 238 TCCATGTACGAGTCTTCTCCATCTCTTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 297  
 Qy 343 -----TTGCCAGGTTCACCATCTGATCAGAAATCAACAAGTGGTCTCCAAAC 384  
 Db 298 GACATCCAGACCTACCTGCGACAGAGTACACCATCTCTCTGTATCAAGAACGACTCC 357  
 Qy 385 GGTGTGTCATCGGTATCTCCAACTTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 444  
 Db 358 GGTGTGTCATCGGTATCTCCAACTTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 417  
 Qy 445 TCCGAGCAGTCCATCAACTTCTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 501  
 Db 418 AAGTCCAAAGTCCATCTTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 477  
 Qy 502 AAGTGTGTTCTGTCACCGTCAACCAATCAACAAGTGGTCTCCAAAC 561  
 Db 478 AAGTGTGTTCTGTCACCGTCAACCAATCAACAAGTGGTCTCCAAAC 537  
 Qy 562 GGTAAGCTGATCGACACCATCAAGGTTCAGCGGTATCAAGTGGTCTCCAAAC 621  
 Db 538 GGTTCCTGTAAGAGTCCGAGAGATCTCAACCTGGAGTCAACCTCCCAACGAC 597  
 Qy 622 ATCACTTCGAGATCAACAGATCCCGACACCGGTCTGATCAGCTCCGACGAC 681  
 Db 598 ATCACTTCGAGATCAACAGATCCCGACACCGGTCTGATCAGCTCCGACGAC 633  
 Qy 682 ATCACTTCGAGATCAACAGATCCCGACACCGGTCTGATCAGCTCCGACGAC 741  
 Db 634 AAGTGTGTTCTGTCACCGTCAACCAATCAACAAGTGGTCTCCAAAC 693  
 Qy 742 AACATCTGTTCACTCTTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 801  
 Db 694 TCCCTCCCTGATCGACATCTCTGATCAGAAATCAACAAGTGGTCTCCAAAC 753  
 Qy 802 CTGAGATCAACAGGAGTACTACATGTTCAACATCGACTTCAAGATGATGATGAC 861  
 Db 754 CTGGTTACGACCCAGTACTACCTGTTCAACAGGATGATGATGATGATGATGAC 813







QY	582	ATCAACATGTGGATCGGTGACCTTACATCTTCGCCAAGGAGTGTGGACGATTAAGACATC	741
Db	643	AAGTCTGTGGATCAAGGACTTCAACATCTTCGGTCTGTGAGCTGAACGCCACGAGGTC	702
QY	742	AACATCTCTTCAACTCCCTTGCAGTACACCAACGTCGTCGAAGGACTACTGGGTAACGAC	801
Db	703	TCCTCCCTGTACTGGATCCAGTCCCTCCACCAACCCCTGAAGGACTTCTGGGAAACCCA	762
QY	802	CTGAGATACAACAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGTAC	861
Db	763	CTCGGTTACGACACCCAGTACTACTGTTCACCCAGGATATGCAGAACATCTACATCAAG	822
QY	862	GCCAACTCCA--GACAGATCGTCTTTCAACACCCAGACGATACAAACGACTTCAACGAG	918
Db	823	TACTTCTCCAAGGCCCTCCATGGGTGAGACCGGCCCTCGTCACTTCAACAACGCCGCC	882
QY	919	GGTTACAAGATCATCATCAAGCGTATCAGAGGTATACCAACCAACGACACCCAGATCAGAGT	978
Db	883	ATCACTTACCAAGACTGTACTCTGGGTCTGGGTTTCATCATCAGAAGGCCTCCACTCC	942
QY	979	GGTGACATCCTGTACTTCGACATGACTATCAACAACAAGGCGCTACAACCTGTTTCATGAAG	1038
Db	943	CGTAACATCAACAACGACACATCGTCCGTGAGGTTGACTACATCTACCTGAACATCGAC	1002
QY	1039	AACGAGACCATGTACCGGACAAACCACTCCACCGAGGAGCATCTAGGCCATCGGTCGCT	1098
Db	1003	ACATCTCCGACGATCTTACCGTGTCTACGTCTCTGGTCTCACTCCAAGGATCCAGACC	1062
QY	1099	GAGCGACCAAGGACATCAACGACACATCATCTTCCAGATCCACGCCAATGAACAACACT	1158
Db	1063	CAGCTGTCTCTGGCCCCAATCAACGACGACCCCTACCTTCTACGAGTCTCGCAGATCAAG	1122
QY	1159	TACTACTACGCTTCCCGAGATCTTCAAGTCCCACTTCAACGGTGAGAACATCTCCGGTATC	1218
Db	1123	AAGTACTACGAGAGCCACTACACACTGCAGATCTCTGTCGAGAGGACACCAAGACC	1182
QY	1219	T 1219	
Db	1183	T 1183	
RESULT 10			
AA54490			
ID	AA54490 standard; DNA; 1317 BP.		
AC	AA54490;		
XX			
DT	11-APR-2001 (first entry)		
DE	Botulism toxin heavy chain C-terminal coding sequence (serotype F).		
KW	Botulism; toxin; neurotoxin; heavy chain; recombinant expression;		
KW	recombinant vector; antigen; Immune response; vaccine; Bacterium;		
KW	infection; ds.		
XX	Synthetic.		
OS	Clostridium botulinum.		
XX			
FH	Location/Qualifiers		
FT	10..1308		
FT	/*tag= a		
FT	/product= H_C peptide fragment		
XX			
PN	WO200067700-A2.		
XX			
PD	16-NOV-2000.		
XX			
PF	12-MAY-2000; 2000WO-US12890.		
XX			
PR	12-MAY-1999; 99US-0133865.		
PR	12-MAY-1999; 99US-0133866.		
PR	12-MAY-1999; 99US-0133867.		
PR	12-MAY-1999; 99US-0133868.		

PR 12-MAY-1999; 99US-0133869.  
PR 28-MAY-1999; 99US-0133873.  
PR 29-JUL-1999; 99US-0146192.  
XX  
XX  
PA ((USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
XX  
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
PI  
DR WFI: 2001-018048/02.  
XX  
DR P-PSDB; AAB04096.  
XX  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulism  
XX  
XX Claim 2; Fig 9a; 73pp; English.  
PS  
XX Botulin neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.  
XX  
XX Sequence 1317 BP: 396 A; 376 C; 239 G; 306 T; 0 other;

Query Match	14.68;	Score	199.8;	DB	22;	Length	1317;
Best Local Similarity	56.7%;	0;	Mismatches	5.1e-34;			
Matches	497;	Conservative	Pre	Indels	48;	Gaps	
Qy	27	CATCTTCTCCTACACCAACAACCTCCCTGGTTGTGAAGGACATCATCAACGAGTAGTACTTCAACAA	86				
Db	6	CACGATGTCTTACACCAACGACAAGATCCTGTACTTGTACTTCAACAAGCTGTACAAGAA	65				
Qy	87	CATCAACGACTCCAGATCTCTGTCCCTGAGAACCGTAGACACACTTGGTCGACACCTC	146				
Db	66	GATCAAGGACAACCTCCATCTTTGGACATGAGATACGAAAAACAATAAGTTTCATCGACATCTC	125				
Qy	147	CGGTTTACAACGCCGAGGTCCTCCGAGGAGGTGACGTCCCACTGAACCCAAATCTTCCCAATT	206				
Db	126	CGGTTACGGTTCACAACATCTCCATCAACGGTGAGCTGTACATCTACTCCACCAATAGAAA	185				
Qy	207	CGACTTCAAGCTGGGTTCTCTCCGGTGAGGACAGAGTAGGTATCTGTCACCCAGACGA	266				
Db	186	CCAGTTTCCGAAATCTACTCTCTCC-----AGCCCTCCGAGGTCAACATCGCTCAGAACA	239				
Qy	267	GAACATCTGCTACAACTCCATGTACGAGTCCCTTCTCCATCTCCCTTCTGGATCAGATCAA	326				
Db	240	CGACATCATCTACAACGGAAGATACCAGAATCTCCATCTCCTTCTGGGTCCTGATATCC	299				
Qy	327	CAAGTGGGTCTCCACTT-----GCCAGGTTACACCATCATCGACTCCGCTC---AA	374				
Db	300	AAAGTACTTTCACAAGGTCACACTGTAATACGAGTAGTACACCATCATCTGACTGCATCCGTAA	359				
Qy	375	GAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTCTTCAACCCCTGAAGCA	434				
Db	360	CAATAACTCCGGATGGAAGATCTCCCTGTAATACACAAGATCATCTGGACCTCTGCAGGA	419				
Qy	435	GAACGAGGACTCCGAGAGTCCATCACTTCTCCTTAGGACATCTCCACACAGCTCCTGG	494				
Db	420	CACCGCGGTAAACAATCAGAGTTGGTCTTCTAACTACACCCAGATGATCTCCATCTCCGA	479				
Qy	495	TTA---CAACAAGTGGTCTCTCGCCACCCGTCACCAACAACATGATGGGTACATGAAGAT	551				

```

Db 480 CTACATCAACAGTGGATCTCGTCACCATCACCAATGCGTTGGGAACCTCCAGAT 539
QY 552 CTACATCAACGGTAAAGTGTGATCGACACCATCAAGGTCAGAGTGTACCGGTATCAACT 611
Db 540 CTACATCAACGGTAAAGTGTGATCGACGAGAGAGTCCATCTCCAAGTGTGGTGACATCCACGT 599
QY 612 CTCGAAGACCATCACTCTCGAGATCAACAAGATCCCGAGACCGGTCTGTATCACCTCCGA 671
Db 600 CTCGACAAACATTTGTTCAAGATCGTGGTGTAAAGACACCCGTT----- 646
QY 672 CTCGACAAACATCAACATGTGATCGGTGATCTTATACATCTTCCCAAGAGTGTGACCG 731
Db 647 -----ACGTCGGATCCGTTACTTCAAAAGTCTTCGACACTGAGTGGTAA 692
QY 732 TAAGACATCAACATCTCTGTTCAACTCTTCAGTACACCAACAGTGTGTCAGGACTACTG 791
Db 693 GACCGAGATCGAGACCTGTGTTACTCCGAGGCGTGCACCCATCCATCTGAGGACTCTG 752
QY 792 GGGTAACGACCTGAGATACACAAGAGGAGTACTATCATGGTCAACATCGACTTGAACAG 851
Db 753 GGGTAACCTGCTGTACACAACAGCTTACTACTTCTGTAACCTTGTGGTACCGACAA 812
QY 852 ATACATGTAGCGCACTCCAGACAGATCGTCTCAAC 888
Db 813 GTCCATCACCCAGAACTCCAACTCTTTGAACATCAAC 849

RESULT 11
AR287216
ID AA287216 standard; DNA; 1317 BP.
XX
AC AA287216;
XX
DT 08-MAY-2000 (first entry)
XX
DE DNA encoding synthetic BoNT serotype F (BoNTF) Hc fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
XX
FH Key
FT Location/Qualifiers
FT 10..1308
FT /product= a
FT /tag= a
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
XX
PR 12-MAY-1999; 99US-0133870.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
XX
DR P-PSDB; AAY7138.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
PS Disclosure; Page 45-46; 54pp; English.
XX

```

```

CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AA287212-287217 represent
CC synthetic DNA sequences encoding BoNT Hc fragments used in the present
CC invention. These were optimised for codon usage for expression in yeast.
XX
SQ Sequence 1317 BP; 397 A; 375 C; 239 G; 305 T; 1 other;

```

```

Query Match 14.5%; Score 198.2; DB 21; Length 1317;
Best Local Similarity 56.6%; Pred. No. 1.1e-33;
Matches 496; Conservative 0; Mismatches 333; Indels 48; Gaps 5;

QY 27 CATCTTCTCTACACCAACCTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAA 86
Db 6 CACGATGCTCTACACCAACGACAGATCTTGATCTGTACTTCAACAAGCTGTACAAGAA 65
QY 87 CATCAACGACTCCAAAGATCTCTGTCCTGACAGACCGGTGAAGAACACCTTGGTGCACACTC 146
Db 66 GATCAAGGACAACTCCATCTTGGACATGAGATACGAAACAATAAGTTTCATCGACATCTC 125
QY 147 CGGTACAAACCGCGAGGTCTCCGAGGAGGTGACGTCCAGCTGAACCCATCTTCCCAT 206
Db 136 CGGTACGGTTCCAACATCTCCATCAACGGTGACGTCTACATCTATCCACCAATAGAAA 185
QY 207 CGACTTAAGTGGTGGTTCCTCGGTGAGGACAGAGTGAAGTCACTCGTCAACCAGAACGA 266
Db 186 CCAGTTCGGAATCTACTCTCTCC-----AAGCCTTCGAGGTCAACATCGCTCAGAACAA 239
QY 267 GAACATGCTGTACAACTCCATGTACGAGTCTCTTCCATCTCTCTGATCAGATCAAGAT 326
Db 240 CGACATCATCTACAAACGAGATACACAGACTTCTCATCTCTTCTGGTCCGTATCCC 299
QY 327 CAAGTGGTCTCCAACTT-----GCCAGGTTACACCATCATCGACTCCGTC---AA 374
Db 300 AAAGTACTTCAACAAGGTCAACCTGAAATAACGAGTACACCATCATCGACTCGATCGTAA 359
QY 375 GAACAACTCCGGTGGTTCCTCATCTCCATCTCCAACTTCTCTGATCTTCACTCCGTGAAGCA 434
Db 360 CAATAACTCCGGATGGAAGATCTCCCTGACTACAGAGATCATCTGGAACCTCGAGGA 419
QY 435 GAACGAGGACTCCGAGCAGTCCATCAACTTCTCTAG---GACATCTCCAAACACCTCC 491
Db 420 CACGCGCGGTAAACAATCAGAAGTGGTCTTCACTACACNAGATGATCTCCATCTCCGA 479
QY 492 TGGTTACAAACAGTGGTCTCTCGTCACCGTCAACCAACAAATGATGGTAAACATGAAGAT 551
Db 480 CTACATCAACAAGTGGATCTTTCGTCACCATCAACAAATGACCGTTGGGAAACATCCAGAT 539
QY 552 CTACATCAACGGTAAAGTGTGATCGACACCATCAAGGTCAAGGAGTGTGACCGGTATCAACT 611
Db 540 CTACATCAACGGTAAAGTGTGATCGACGAGAGTCCATCTCCAACTTGGGTGACATCCACGT 599
QY 612 CTCGAAGACCATCACTCTCGAGATCAACAAGATCCCGAGACCGGTCTGTATCACCTCCGA 671
Db 600 CTCGACAAACATTTGTTTCAAGATCGTGGTGTGAACGACACCCGTT----- 646
QY 672 CTCGACAAACATCAACATGTGGATCCGTCGACTTCTATCATCTTCCCAAGAGTGTGACCG 731

```

```

Db 647 -----ACGTCGGATCCGTTACTTCAAGTCTTCGACACGTAGTTGGTAA 692
Qy 732 TAAGGACATCAACATCTGTTCAACTCCTTTCAGTACACCAACGCTGCTCAAGGACTACTG 791
Db 693 GACCGAGATCGAGACCTTGTTACTCCGACGAGCTGACCCATCCATCCTGAAGACTTCTG 752
Qy 792 GGGTAACGACCTGAGATACAAACAGGNGTACTACATGCTCAACATCGACTACTTGAACAG 851
Db 753 GGGTAACATCCTGCTGTACAAACAGGTTACTACTTGTGAAACCTGTTGCGTACCGACAA 812
Qy 852 ATACATGATAGCCCAACTCCACAGACAGATCGTCTTCAAC 888
Db 813 GTCATCAACCCAGAACTCCAACTCTTGAACATCAAC 849

RESULT 12
AAV30576
ID AAV30576 standard; DNA; 1351 BP.
XX
AC AAV30576;
XX
DI 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in pHisBotA(syn).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism; ds.
XX
OS Clostridium botulinum serotype A.
OS Synthetic.
XX
FH Key
FT Location/Qualifiers
FT 1..1338
FT /*tag= a
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Thalley BS, Williams JA;
XX
P-PSDB: AAW68391.
XX
WPI: 1998-230234/20.
XX
Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
Example 29; Page 279-281; 428pp; English.
XX
This is the DNA sequence of the Clostridium botulinum serotype A
CC toxin C fragment gene contained in plasmid pHisBotA(syn). The
CC encoded toxin A polypeptide (see AAW68391) has a histidine-tagged
CC N-terminal extension. The vector was used to express native
CC (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins. Methods are provided which allow for the
CC isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells.
CC The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 1351 BP; 407 A; 348 C; 247 G; 349 T; 0 other;

```

```

Query Match 12.4%; Score 169.6; DB 19; Length 1351;
Best Local Similarity 54.4%; Pred. No. 28-27; Indels 42; Gaps 4;
Matches 455; Conservative 0; Mismatches 339;
Qy 44 ACAACTCCCTGTTGAAGGACATCATCAAGAGTACTTCAACAACTCAACGACTCCAAGA 103
Db 20 ACATGGCTCCTCTGCTGTACCTTCACTGAATACATCAAGAATCATCAATCACTCCA 79
Qy 104 TCCGTCCCTGCGAGACCGTAAAGAACACCTTGGTCACACCTCCGGTTACAACGCGGAG 163
Db 80 TCCGTGAACCTGCGCTACGAATCAATCACTGATCGCTGTCTCGTACGTTCCAAA 139
Qy 164 TCTCCGAGGAGGTGAGCTCCAGCTGAACCCAAATCTCCCATTCGACTTCAAGCTGGGT 223
Db 140 TCAACATCGGTTCTAAGTTAACTTCGATCCGATCCGAAGATCAGATCCAGCTG--- 195
Qy 224 CTTCCGGTGAGGACAGAGTAAAGTTCATGTCACCCAGAACGAGAACATCGTCTCAACT 283
Db 196 --TTCAATCTGGAATCTTCCAAAATCGAAGTTATCTCTGAAGATGCTATCGTATACA 253
Qy 284 CCATGTAGAGTCTCTTCCATCTCTTGGATCAGATCAACAGTGGTGGTCTCCAAC 343
Db 254 CTAIGTACGAAAACCTCTCCACCTCTTCTGGATCGTATCCGGAATFACTTCAACTCCA 313
Qy 344 TGCC-----AGTTACACCATCATCGACTCCGTCAGTCAAGAACAACTCCGTTTGGTCCA 394
Db 314 TCTCTCTGAACAATGAATACACCATCATCACTGATGGAAGAAACAATCTGGTTGGAAAG 373
Qy 395 TCGGTATCATCTCCAACCTCTCTGCTTCACTGAGAGAGAGGAGGACTCCGAGCAGT 454
Db 374 TATCTCTGAAGTACGGTGAATCATCTGGACTCTGAGGACACTCAGGAATCAACAGC 433
Qy 455 CCATCAACTCTCTCTACGACATCTCCAACAACCTCTCTGTTTACACAA---GTGGTTCT 511
Db 434 GTGTTGATTCAAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGTGGATCT 493
Qy 512 TCGTACCGCTCACCACACATGATGGTAAATCAATGAATCTACATCAACGCTAAGCTGA 571
Db 494 TCGTTACCATCACCACAAATCGTCTGAATAAATCCAAAATCTACATCAACGGCCGCTGA 553
Qy 572 TCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAGACCATCACTTCG 631
Db 554 TCGACCAAGAACCGATCTCCAATCTGGTAAACATCCACGCTTCTAATAACATCATCTTCA 613
Qy 632 AGATCAACAAGATCCACAGACCGGTCTGATCACTCCGACTCCGACACATCAACATGT 691
Db 614 AACTGGACGGTTCTGTCGACACTC-----ACCGTCAATCT 649
Qy 692 GGATCCGTGACTTCTACATCTTCGCAAGAGTTGGACGGTAAAGGACATCAACATCCTGT 751
Db 650 GGATCAAAATCTTCAATCTGTTCCGACAAAGACTGAACGAAAAGAAATCAAAGACCTGT 709
Qy 752 TCAACTCCTTGCAGTACACCAAGTGTGTCAGGACTACTGGGTACGACCTTGAGATACA 811
Db 710 ACGAACACCACTGCTCAATCTCTGTTATCTCTGAAGACTCTCTGGGTGACTTACCTGCA 769
Qy 812 ACAAGGAGTACTACATGTTCAACATCGACTTGTGAACAGATACATGTAGCGCAAC 867
Db 770 ACAAACGTTACTACTGCTGAATCTGTACGATCCGACAAATACGTTGAGGTCAAC 825

RESULT 13
AAAS4483
ID AAAS4483 standard; DNA; 1323 BP.
XX
AC AAAS4483;
XX
DI 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal coding sequence (serotype A).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;

```



PR 14-APR-1995; 95US-0422711.  
XX (OPHI-) OPHIDIAN PHARM INC.  
XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
XX Williams JA;  
XX WPI: 1996-230603/23.  
DR P-PSDB: AAR95009.  
XX  
XX Fusion proteins comprising non-toxin protein and part of toxin  
XX useful to form anti-toxins against Clostridium botulinum type A, and  
XX C. difficile type toxins, and to treat C. difficile intoxication,  
XX partic. diarrhoea  
XX Example 24; Page 340-342; 434pp; English.  
XX  
XX A nucleotide sequence (AAT29246) present in vector pETHisa encodes  
XX the phisBot fusion protein (AAR95009) comprising a polyhistidine  
XX affinity tag and fragment C (see also AAR95008) of the Clostridium  
XX botulinum type A neurotoxin. The phisBot protein was expressed  
XX in Escherichia coli as a soluble protein and was purified by  
XX metal chelate affinity chromatography to obtain a product free  
XX of endotoxin contamination that may be useful as an immunogen  
XX in vaccine compns.  
XX  
XX Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;  
SQ

Query Match 12.4%; Score 169.4; DB 17; Length 1402;  
Best Local Similarity 54.4%; Pred. No. 2.2e-27;  
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;

QY 41 CCAACAACCTCCCTGTTGAAGGACATCATCAAGAGACTTCAACAACATCAACGACTCCA 100  
DB 68 CTAGCATGGCTCGTCTGCTGTCACCTTCACTGAATACATCAAGAACATCATCACTAC 127  
QY 101 AGATCTCTGCTCCGACCAACGGTAGACACCTTGTGACACCTCCGGTTACAGCGCG 160  
DB 128 CCATCTCTGAACCTCGGTACGAAATCCAAATCACTGATCGACCTGTCTCGCTACGGTTCCA 187  
QY 161 AGGCTCCGAGGAGGGTGAAGTCCAGCTGAACCAATCTCCCATTCGACTTCAAGCTGG 220  
DB 188 AATACACATCGGTTCTTAAGTTAACTTCGATCGGATCGACAGATCATGATCCAGCTG- 246  
QY 221 GTTCTCGGTGAGGAGAGGTAAAGTCACTGTCACCCAGACGAGAACATCGTCTACA 280  
DB 247 -----TTCAATCTGGAATCTTCCAAATCGAAGTTATCTCTGAAAGATGCTATCGTATACA 301  
QY 281 ACTCATGTAGAGTCTCTTCCATCTCCTTGTGGATCAGAAATCAACAAGTGGTCTCCA 340  
DB 302 ACTCTATGTAGAAACTTCTCCACCTCTCTGGAATCCGATCCCGAATATCTCACTCACT 361  
QY 341 ACTTGCC-----AGGTTACACCATCATCATCGACTCCGGTCAAGAACACTCCGGTTGTT 391  
DB 362 CCATCTCTGGAACAATGAATACCAATCATCACTCACTGATGCGATGGAACAATCTCTGTTGA 421  
QY 392 CCATCGGTATCATCTCCAATCTCTGCTCTCACTGAGTGAAGAGAGGAGTCCGAGC 451  
DB 422 AAGTATCTCTGAATAGGTGAATATCTCTGAGTCTGCGAGGACATCGAGAAATCAAC 481  
QY 452 AGTCCATCACTTCTCTAGACATCTCCAAACAGCTCTCTGTTTACAACA-----GTGTT 508  
DB 482 AGCGTGTGTATTCAAAATCTCTCAGATGATCAACATCTCTGACTACATCAATCGGTGA 541  
QY 509 TCTTCGTCACCGTCAACAACAATGATGGTGAATGATGAGATCTACATCAACGGTAAAC 568  
DB 542 TCTTCGTTTACCATCAACAATCTGTTGAATCACTCAAAATCTACATCAACGGCGGTC 601  
QY 569 TGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCACTTCTCCAAAGACATCACT 628  
DB 602 TGATCGACCAAAACCGATCTCCAATCTGGGTACATCCACGCTTCTTAATACATCATGT 661  
QY 629 TCGAGATCAACAAGATCCAGACACCGGTCTGATCACTCCGACTCGGACATCAACA 688

DB 662 TCAAACTGGACGGTGTGTCGACACTC-----ACCGCTACA 697  
QY 689 TGTGGATCCGTCGACTTCTTACATCTTCGCCAAGGAGTGTGACGTAAGGACATCAACATCC 748  
DB 698 TCTGGATCAAAATCTTCAATCTCTGACAAAGAACTGAACGAAAAAGAAATCAAGACC 757  
QY 749 TGTCAACTCCTTGCAGTACACCAACGTCGTCAAGGACTACTGGGTAACGACCTTGAGAT 808  
DB 758 TGTACGACCAACCACTCCAAATCTCTGGTATCTCTGAAAGACTTCTGGGTGACTACCTGCGAGT 817  
QY 809 ACACAAAGGAGTACTACATGTTCAACATCGACTTCTGAACAGATACATGTCAGGCAAC 867  
DB 818 ACACAAACCGTACTACATGCTGAATCTGTAGATCCGAACAATACGTTGACGTCAAC 876

RESULT 15  
AAV30572  
ID AAV30572 standard; DNA; 1402 BP.  
XX  
AC AAV30572;  
XX  
DI 07-DEC-1998 (first entry)  
XX  
DE Clostridium botulinum toxin A fragment C gene in phisBot.  
XX  
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;  
XX botulism; ds.  
XX  
OS Clostridium botulinum serotype A.  
OS Synthetic.  
XX  
FE Key Location/Qualifiers  
FT CDS 1..1389  
FT /\*tag= a  
XX  
XX W09808540-A1.  
XX  
XX 05-MAR-1998.  
XX  
XX 28-AUG-1997; 97WO-US15394.  
XX  
XX 28-AUG-1996; 96US-0704159.  
XX  
XX (OPHI-) OPHIDIAN PHARM INC.  
XX Thalley BS, Williams JA;  
XX  
XX WPI: 1998-230234/20.  
XX P-PSDB: AAW68390.  
XX  
XX Host cell containing recombinant expression vector encoding  
XX Clostridium botulinum type B or E toxin - useful to treat humans  
XX and other animals at risk of intoxication with clostridial toxin  
XX  
XX Example 24; Page 365-367; 428pp; English.  
XX  
XX This is the DNA sequence of the Clostridium botulinum serotype A  
XX toxin C fragment gene contained in plasmid phisBot. The encoded  
XX toxin A polypeptide (see AAW68390) has a histidine-tagged  
XX N-terminal extension. The vector was used to express native  
XX (i.e. non-fusion) soluble C fragment in Escherichia coli host  
XX cells. The invention relates to recombinant proteins derived from  
XX C. botulinum toxins. Methods are provided which allow for the  
XX isolation of soluble recombinant proteins free of significant  
XX endotoxin contamination. Preferred hosts for production of  
XX recombinant proteins are E. coli, insect cells and yeast cells.  
XX The recombinant toxins are used as immunogens for the production  
XX of vaccines and antitoxins that are useful in the treatment of  
XX humans and animals at risk of intoxication with clostridial toxin.  
XX  
XX Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;  
SQ



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:20:20 ; Search time 57 Seconds  
(without alignments)  
7376.383 Million cell updates/sec

Title: US-09-910-186A-9  
Sequence: 1 gaattacagatgaccatcccc.....tctccgagtagaagaattc 1371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/PCTUS-COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	169.4	12.4	1402	1	US-08-480-604A-25
2	169.4	12.4	1402	2	US-08-405-496A-25
3	169.4	12.4	1402	4	US-08-915-136-25
4	188.8	12.3	1330	1	US-08-480-604A-22
5	188.8	12.3	1330	2	US-08-405-496A-22
6	168.8	12.3	1330	4	US-08-915-136-22
7	136	9.9	5163	3	US-08-700-651-1
8	136	9.9	5163	3	US-08-928-361B-4
9	136	9.9	5318	3	US-08-700-651-2
10	136	9.9	5318	3	US-08-928-361B-3
11	115.4	8.4	5511	3	US-08-928-361B-2
12	115.4	8.4	7334	3	US-08-928-361B-1
13	95	6.9	4483	4	US-08-961-527-363
14	93.4	6.8	2004	1	US-08-471-033-18
15	93.4	6.8	2004	2	US-08-471-044-18
16	93.4	6.8	2004	2	US-08-463-483A-18
17	93.4	6.8	2004	2	US-08-471-046A-18
18	93.4	6.8	2004	2	US-08-470-568B-18
19	93.4	6.8	2004	2	US-08-469-334-18
20	93.4	6.8	2004	3	US-09-300-529-18
21	93.4	6.8	2576	1	US-08-471-033-35
22	93.4	6.8	2576	2	US-08-471-044-35
23	93.4	6.8	2576	2	US-08-463-483A-35
24	93.4	6.8	2576	2	US-08-471-046A-35
25	93.4	6.8	2576	2	US-08-470-568B-35
26	93.4	6.8	2576	2	US-08-469-334-35
27	93.4	6.8	2576	3	US-09-300-529-35

28	93.4	6.8	2555	1	US-08-471-033-17
29	93.4	6.8	2555	1	US-08-471-033-26
30	93.4	6.8	2555	2	US-08-471-044-17
31	93.4	6.8	2555	2	US-08-471-044-26
32	93.4	6.8	2555	2	US-08-463-483A-17
33	93.4	6.8	2555	2	US-08-463-483A-26
34	93.4	6.8	2555	2	US-08-471-046A-17
35	93.4	6.8	2555	2	US-08-471-046A-26
36	93.4	6.8	2555	2	US-08-470-568B-26
37	93.4	6.8	2555	2	US-08-470-568B-17
38	93.4	6.8	2555	2	US-08-469-334-17
39	93.4	6.8	2555	2	US-08-469-334-26
40	93.4	6.8	2555	3	US-09-300-529-17
41	93.4	6.8	2555	3	US-09-300-529-26
42	93.4	6.8	4031	1	US-08-471-033-49
43	93.4	6.8	4031	2	US-08-471-044-49
44	93.4	6.8	4031	2	US-08-463-483A-49
45	93.4	6.8	4031	2	US-08-471-046A-49

ALIGNMENTS

RESULT 1

US-08-480-604A-25

Sequence 25, Application US/08480604A

Patent No. 5736139

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHYE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,604A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION NUMBER: US 08/422,711

FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

Query Match 12.4%; Score 169.4; DB 2; Length 1402;  
Best Local Similarity 54.4%; Pred. No. 8e-29;  
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps





```

QY 452 AGTCATCAACTTCTCCACGACATCTCCACACAGCGTCTCGTTACAAACA---GTGGT 508
Db 482 ACCGGTGTGTTATCAATCAATCTCTAGATGATCAACATCTCTGATCAATCAATCGTGG 541
QY 509 TCTTCTGTCACCTGTCACCAACACATGATGGTAACTGATGATCAATCAACGGTAAAGC 568
Db 542 TCTTCTGTCACCTGTCACCAACATGATGGTAACTGATGATCAATCAACGGTAAAGC 601
QY 569 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 628
Db 602 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 661
QY 629 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 688
Db 662 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 697
QY 689 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 748
Db 698 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 757
QY 749 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 808
Db 758 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 817
QY 809 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 867
Db 818 TGATCGACACATCAAGTCAAGGAGTGGCCGATGATCACTCTTCCAGACCATCACT 876

```

RESULT 4

```

US-08-480-604A-22
; Sequence 22, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1314
; US-08-480-604A-22

Query Match 12.3%; Score 168.8; DB 1; Length 1330;
Best Local Similarity 54.8%; Pred. No. 1.1e-28;
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;

QY 64 ATCATCAAGAGTACTTCAACACATCAACGAGTCCCAAGATCCTGCTCCCTGCAAGACCGT 123
Db 19 ACCCTCATGATACATCAAGACATCATCAATACCTCCATCCTGACCTCGCTAGCA 78
QY 124 AAGAACACCTTGGTGCACACCTCCGGTTTACACCGCGAGGTCTCCGAGGAGGTGACGTC 183
Db 79 TCCAAATCACTGATCGACCTGCTCGCTGCTCCGCTTCCAAATCAACATCGGTTCTAAAGT 138
QY 184 CAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGTTCCTCCGCTGAGGACAGAGT 243
Db 139 AACCTCGATCGATCGACAGATCAAGATCCAGCTG-----TTCATCTGGATCTTC 192
QY 244 AAGGTATCTGTCACCCAGAGAGACATCGTCTTACAACTCCATGTCGATGAGTCTCTCC 303
Db 193 AAAATCGAAGTATCTGAGAAATGCTATGATACAACTCTATGTACGAAATCTCTCC 252
QY 304 ATCTCTTCTGATCAGAAATCAACAACTGGTCTCCAACTTGGC-----AGTTAC 354
Db 253 ACCTCTTCTGATCGGATCCCGAATACCTCAACTCCATCTCTGACATGATATAC 312
QY 355 ACCATCATGACTCGGTCAAGACAACTCCGGTTGTCGATCGGTATCTCTCCAACTTC 414
Db 313 ACCATCATCAACTGCATGGAACAAATCTGTTGGAAGTATCTCTGAACTACGTTGA 372
QY 415 CTGGTCTTCCCTGAAGCAGAGGAGTCCGAGGAGTCCGATCACTTCTCTTACGAC 474
Db 373 ATCATCTGACTCTGAGGACACTCAGGAATCAACAGCGGTGTGTATTCATATCTCT 432
QY 475 ATCTCCAAACAGCTCTCTGTTTACAAACA---GTGGTCTTCTGTCGCTCCGTCACCAAC 531
Db 433 CAGATGATCAACATCTCTGACTACATCAATCGCTGATCTTCTGTTACCATCAACAAT 492
QY 532 ATGATGGGTATCATGATCTATCACTCAACGTAAGTCTGATCGACCATCAAGTCAAG 591
Db 493 CGTCTGAATCACTCAAAATCTATCTACGCGGCTGTATCGACGAGAACGATCTCC 552
QY 592 GAGTTGACCGGTATCACTTCTCCAGACCATCACTCCGATCGAGATCAACAGATCCAGAC 651
Db 553 AATCTGGGTAACATCCACGCTTCTTAATAACATCATGTTCAACTGGACGCTTCTGCTG 612
QY 652 ACCGCTCTGATCACTCCGACTCCGACCAACATCAACATGATGAGTCCGCTGATCTTCACT 711
Db 613 ACTC-----ACCGCTACATCTGGATCAAAATCTTCAATCTCTG 648
QY 712 TTCGCAAGAGGATGGAGGTAAGGACATCAACATCTCTGTTCAACTCTTTCAGTACACC 771
Db 649 TTCGCAAGAGGATGGAGGTAAGGAAAGAAATCAAGACCTGTACGACCAACAGTCCATCT 708

```

QY 772 AACGTCGTCAGGACTACTGGGTAACGACCTGAGATACAAACAGGAGTACTACATGGTC 831  
DB 709 GGTATCCTGAAGAGACTTGTGGGTGACTACTGCTGAGTACGACAAACCGTACTACATGCTG 768  
QY 832 AACATCGACTACTTGAACAGATACATATGACGCAAC 867  
DB 769 AATCTGTACGATCCGACAAATAGTTGACGTCAAC 804

## RESULT 5

US-08-405-496A-22  
; Sequence 22, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1330 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: Gdouble  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1314  
US-08-405-496A-22

Query Match 12.39; Score 168.8; DB 2; Length 1330;  
Best Local Similarity 54.8%; Pred. No. 1.1e-28;  
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;  
QY 64 ATCATCAACGAGTACTCAACAACTCAACGACTCCAGATCTGTCCTGCGAGAACCGT 123  
DB 19 ACCCTTCACTGATACATCAAGACATCATCAATCTCCATCTGACCTGCGCTACGAA 78

QY 124 AAGAACACCTTGGTTCGACACCTCGGGTTACACCGGAGGTCTCCGAGGAGGTTGACGTC 183  
DB 79 TCCAACTCACCTGTGATCGACCTGTCTCGTACGCTTCCAAAATCAACATCGTTCTTAAAGTT 138  
QY 184 CAGTGAACCAACTCTTCCCATTCGACTTCAAGCTGGGTTCCCTCCGGTGGAGACAGAGT 243  
DB 139 AACTTCGATCGATCGACAAAGATCAGATCCAGCTG-----TTCAATCTGGATCTTCC 192  
QY 244 AAGTTCATCGTCAACCCAGAGAACATCGTCTACAACTCCATCCATGAGTCTCTTCTCC 303  
DB 193 AAAATCGAAGTTATCTCTGGAAGATGCTATCGTATACAACTCTATGACGAAAATCTCTCC 252  
QY 304 ATCTCCTCTGGATCAGAATCAACAGTGGTCTCCAACTTGC-----AGGTTAC 354  
DB 253 ACCCTCTCTGGATCCGATATCCCGAATATCTCACTCCATCTCTCTGACAAATGAATAC 312  
QY 355 ACCATCATCGACTCCGTCAGAGAACACTCCGGTGGTGGTCCATCGTATCATCTCCAAATTC 414  
DB 313 ACCATCATCAACTGCATGGAAGAACATCTGTTGGAAAGTATCTCTGAATACGAGTAA 372  
QY 415 CTGCTCTTCAACCTGAGCAGAGAGGACTCCGAGAGTCCGAGCAGTCCATCAACTCTCTTACGAC 474  
DB 373 ATCATCTGGACTCTGAGGACACTCAGGAATCAACAGCGGTGTTGATTCAAATACTCT 432  
QY 475 ATCTCCAACACGCTCTCTGGTTTACAA-----GTGGTTCTTCTCTACCTCCACCAACAC 531  
DB 433 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCTGTTACCATCACCACAT 492  
QY 532 ATGATGGTAACTGAAGATCTACATCAACGGTAAAGCTGATCGACACCATCAAGTCAAG 591  
DB 493 CGTCTGAATAACTCCAAAATCTATCATCAGCGCGCTCTGATCGACGAGAAACCGATCTCC 552  
QY 592 GAGTTGACCGGTATCAACTTCTCCAAGACCATCACCTTTCAGATCAACAGATCCAGAC 651  
DB 553 AATCTGGTAACTCCAGCTTCTTAATCAATCATGTTCAAACTGGACGCTGTCGTCGAC 612  
QY 652 ACCGGTCTGATCACTCCGACTCCGACACATCAATCAAGATGAGTGGATCGTACTCTACATC 711  
DB 613 ACTC-----ACCGTACATCTGGAATCAAAATCAATCTG 648  
QY 712 TTCGCAAGGAGTTGGACGTTAGGACATCAACATCTCTTCAACTCTTGGAGTACAC 771  
DB 649 TTCGCAAGGAGTGAAGGAAAGAAATCAAGACCTGTACGACAAACAGTCCATCT 708  
QY 772 AACCTCTGAGGACTACTGGGTGAGCTGAGTACAAAGAGGAGTACTACATGTC 831  
DB 709 GGTATCCTGAAAGACTTCTGGGTGACTACTGCTGAGTACGACAAACCGTACTACATGTC 768  
QY 832 AACATCGACTACTTGAACAGATACATGTCAGCAAC 867  
DB 769 AATCTGTACGATCCGAAATAGTTGACGTCAAC 804

## RESULT 6

US-08-915-136-22  
; Sequence 22, Application US/08915136  
; Patent No. 5290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTI-TOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104  
 COMPUTER TYPE: FLOPPY disk  
 MEDIUM TYPE: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA: US/08/915,136  
 FILING DATE: 02-DEC-1993  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA: US/08/405,496  
 FILING DATE: 16-MAR-1995  
 PRIOR APPLICATION DATA: US/08/480,604  
 FILING DATE: 04-DEC-1992  
 PRIOR APPLICATION DATA: US/07/985,321  
 FILING DATE: 04-DEC-1992  
 PRIOR APPLICATION DATA: US/07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: INGOLIA, DIANE E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPHD-01763  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1330 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1314  
 US-08-915-136-22

Query Match 12.3%; Score 168.8; DB 4; Length 1330;  
 Best Local Similarity 54.8%; Pred. No. 1.1e-28;  
 Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;  
 QY 64 ATCATCAGGAGTACTTCAACAACATCAACGACTCCAGATCCTGTCCTCCGAGACCGT 123  
 DB 19 ACCTTCACTGAATCATCAAGAACATCATCAATACCTCCATCTGGAACCTGCGCTAGAA 78  
 QY 124 AAGAACACCTTGGTCGACACCTCCGGTTACAAACGCGAGGTCTCCGAGAGGGGTGAGTC 183  
 DB 79 TCCAAATCACTGATCGACCTGTCTCGTACGCTTCCAAATCAACATCGTTCATAGTT 138  
 QY 184 CAGCTGAACCAATCTTCCGATTCGACTTCAAGCTGGTTCCTCCGGTGGAGACAGGT 243  
 DB 139 AACTTCGATCCGATCGAAGAATCAGATCAGCTG-----TTCAATCTGGAATCTTC 192  
 QY 244 AAGGTTCATCGTCAACCCAGAGAACATCGCTTACAACTCCATGTAGGATCCTCTCC 303  
 DB 193 AATATCGAAGTATCTCTGAGAGTCTATCTATACACTCTATGTAGCAAACTTCTCC 252  
 QY 304 ATCTCTTCTGGATCAGATCAACAAAGTGGGTCTCCAACTTGC-----AGGTAC 354  
 DB 253 ACCTCTTCTGGATCGGTATCCGGAATACCTCAACTCCATCTCTGTGAACAATGATAC 312  
 QY 355 ACCATCATCGACTCGGTCAAGAACAACTCCGGTGTGTCATCGGTATCATCTCCAACTTC 414

Db 313 ACCATCATCAACTGCATGGAAACAATCTGGTTGGAAGTATCTCTGAACATCGGTGA 372  
 QY 415 CTGGTCTTCCCTTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTTACGAC 474  
 Db 373 ATCATCTGGACTCTGCGAGACACTCAGGAAATCAAAACAGCGTGTGTATTAATAACTCT 432  
 QY 475 ATCTCCAAACACGCTCTGGTTACAAACA---GTGGTCTTCTGTCACGCTCACCACACAC 531  
 Db 433 CAGATGATCAACATCTCTGACTATCAATCGCTGGATCTCTGTACCATCAACAAT 492  
 QY 532 ATGATGGGTAAACATGAAGATCTATCAACGCGTAAGTGTATCGACACCATCAAGGTCAAG 591  
 Db 493 CGTCTGAATAACTCCAAATCTATCAACGCGCTGTATCGACACCAAGAACGATCTCC 552  
 QY 592 GAGTGGACCGGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAAGATCCCGAC 651  
 Db 553 AATCTGGGTAACTCCACGCTTCTAATAACATCATGTCAAACTGGACGGTGTGCTGAC 612  
 QY 652 ACCGCTGTGATCACTCCGACTCCGACATCAACATCAACATGTGGATCGGTGACTTCTACATC 711  
 Db 613 ACTC-----ACCGCTACATCTGGATCAATACTTCAATCTG 648  
 QY 712 TTCGCCAAGAGTTGGAGCGTAAGGACATCAACATCTCTGTCAACTCTTGCAGTACAC 771  
 Db 649 TTCGACAAAGAACTGAACGAAAGAAATCAAGACCTGTACGACCAACCGTCCCAATCT 708  
 QY 772 AACGTCGTCGAAGGACTACTGGGGTAAAGCCTGAGATACAAACAAGGAGTACTACTGTC 831  
 Db 709 GGTATCCTGAAAGACTTCTGGGTGACTCTCGAGTACGACAAACCGTACTACTGTCG 768  
 QY 832 AACATCGACTACTTGAACAGATCATATGATAGCGCAAC 867  
 Db 769 AATCTGTACGATCGAACAATAATAGTTGACGTCAAC 804  
 RESULT 7  
 US-08-700-651-1  
 ; Sequence 1, Application US/08700651B  
 ; Patent No. 6015882  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PETERSEN, CAROLYN  
 ; APPLICANT: LEECH, JAMES  
 ; APPLICANT: NELSON, RICHARD, C.  
 ; APPLICANT: GUT, JIRI  
 ; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
 ; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
 ; TITLE OF INVENTION: INFECTIONS  
 ; FILE REFERENCE: 480.19-4(HV)  
 ; CURRENT APPLICATION NUMBER: US/08/700,651B  
 ; EARLIER FILING DATE: 1997-08-14  
 ; EARLIER FILING DATE: 1995-04-03  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 5163  
 ; TYPE: DNA  
 ; ORGANISM: Cryptosporidium parvum  
 US-08-700-651-1  
 Query Match 9.9%; Score 136; DB 3; Length 5163;  
 Best Local Similarity 50.6%; Pred. No. 2.9e-21;  
 Matches 328; Conservative 0; Mismatches 320; Indels 0; Gaps 0;  
 QY 541 AACATGAAGATCTATCAACGCGTAAGTGTATCGACACCATCAAGGTCAAGGAGGTGACC 600  
 Db 577 AACACACACACACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 636  
 QY 601 GGTATCAACTTCTCCAGACCATCACTTTCGAGATCAACAAGATCCCAACACACCGGTCTG 660  
 Db 637 AACAAACAACTACTACAACTACTACCACTACTACTACTACTACTACTACTACTACTACTACT 696  
 QY 661 ATCACTCCGACTCCGACAAACATCAACATGTGGATCGGTGACTTCTTACTACTTCTCGCCAAG 720







```

; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-1

Query Match      8.4%; Score 115.4; DB 3; Length 7334;
Best Local Similarity 49.0%; Pred. No. 1.2e-16;
Matches 308; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 606 CAATCTCCAGACATCACTTCGAGATCAACAGATCCAGACACCGGTCTGATCAC 665
DB 2621 CTACTACTAGCAACAAACAAACAAACAAACAAACAAACAACTACTACTACTACCA 2680

QY 666 CTCGACTCCGACAAACATCAACATGTGGATCGGTGACTTCTACATCTTCGCCAAGGAGTT 725
DB 2681 CTACTACTAGCAACAAACAACTACTACTAGCAACAAACAAACAACTACTACTACTA 2740

QY 726 GGACGGTAAGGACATCAATCCTGTTCAATCCTTGTGAGTACACCAACGTCGTCAGGA 785
DB 2741 CTACACCAACAACTCAACCAACCAACCAACCAACCAACCAACCACTACCAAGAAC 2800

QY 786 CTACTGGGTAAAGCACTGAGATCAACAAAGGAGTACTACTGTTCAACATCGACTACTT 845
DB 2801 CAACAAACAAACAAACAAACAACTACTACTACTACTACTACTACTACTACTACTA 2860

QY 846 GACAGATACATGTAGCGCACTCAGACAGATCGTCTTCAACCCAGACGTACACAA 905
DB 2861 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 2920

QY 906 CGACTCAACGAGGTTTACAGATCATCATCAAGCGTATCAGAGGTAACACCAACGACAC 965
DB 2921 CCACACTACCAAGAAACAAACAAACAAACAAACAAACAAACAACTACTACTA 2980

QY 966 CAGAGTCAGAGGTGGTACATCCTGTACTTTCACATGACTATCAACAAACAGGCTTCAA 1025
DB 2981 CCACGAAACAAACAAACGAGCAACCAACCAACCAACCACTACTACTACTACTACTA 3040

QY 1026 CCGTTCATGAAGACGAGACATGTACCGGACACACCACTCCACCGAGGACATCTAGC 1085
DB 3041 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 3100

QY 1086 CAGTGGTCTGCGTGACGACCAAGGACATCAACGACAAACATCATCTTCCAGTCCAGCC 1145
DB 3101 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 3160

QY 1146 AATGAACACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAA 1205
DB 3161 CACACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3220

QY 1206 CACTCCGGTATCTGTTCCATCGGTACCT 1234
DB 3221 CCACAAACAACTACTACTTCTGAACT 3249

RESULT 13
US-08-961-527-363
; Sequence 363, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/961.527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 363:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-363

Query Match      6.9%; Score 95; DB 4; Length 4483;
Best Local Similarity 44.7%; Pred. No. 3.4e-12;
Matches 558; Conservative 0; Mismatches 665; Indels 24; Gaps 4;

QY 29 TCTTCTCTTACACCAACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAACA 88
DB 2362 TCAGGCTCTGAATCGCATCAACGAGTGGTCAGCCTCAGCAAGTACTAGTGCATCAGCA 2421

QY 89 TCACGACTCCAGATCTCTGTCCTTGAGAACCGGTAGAACACCTTGGTCGACACCTCG 148
DB 2422 TCAGCATCAACGAGTGCATCGGCTTCAGCAAGTACCAAGCGCTCAGCTTCAGCAAGCACC 2481

QY 149 GTTACAAACCCGAGGTCTCCGAGAGGAGGTGAGCTTCCAGCTGAACCCCAATCTTCCCATTCG 208
DB 2482 AGTGGCTCAGCTCAGCAAGTACCAAGCGCTCAGCCTCAGCAAGCACCAGTGCCTCAGCT 2541

QY 209 ACTTCAAGCTGGGTTCTCCGGTGAGGACAGAGGTAGGTATCGTACCCAGAACGAGA 268
DB 2542 TCAGCAAGTACCAAGTGGCTCAGCCTCAGCGTGAGCAAGTGGCTTTCAGCAAGTACC 2601

QY 269 ACATCGCTTACAACTCCATGTACGAGTCTTCTCCATCTCCTTCTGATCAGATCAACA 328
DB 2602 TCAGCTCTGATCAGCATCAACGAGTGCATCAGCTTCAGCATCAACAGTGTCTCAGCT 2661

QY 329 AGTGGGTCTCCAACTTCCAGGTTACACCATCATCGATCGGTCCGTCAGCAAGTCCGGTT 388
DB 2662 TCAGCAAGTACCAAGTGGCTCAGCCTCAGCATCAACGAGTGTCTCAGTCTCAGCGTCAACC 2721

QY 389 GGTCCATCGGTAT---CATCTCCAACTTCTTCCCTTCCCTTCAAGCAAGCAAGGAGCT 445
DB 2722 AGTGGCTCTGATCAGCATCAACAGTGCCTCGGCTTCAGCAAGCACCAGTGCCTCGGCT 2781

QY 446 CCGAGCAGTCCATCAACTTCTCTAGACATCTCCAAACAGCTCCCTGGTTACAAACAGT 505
DB 2782 TCAGCAAGTACTAGTGCATCGGCTTCAGCATCGACAGTGGCTGTGAATCGGCTCAACG 2841

QY 506 GGTCTCTCTGCTACCGTTCACCAACA-----CATGATGGTAAACATGAAGATCTACA 556
DB 2842 AGTGGCTTCCGCTTACGATCAACAGTGCCTCGGCTTCAGCAAGCACCAGTGTCTGAA 2901

QY 557 TCACCGGTAAAGTGTATCGACACCATCAAGGTCAAGGATTTGACCGGTATCAACTCTTCCA 616
DB 2902 TCTGCTCAACCAAGTGGTCCGCTTCAGCGTCAACAGTGGCTTTCAGCGTTCGACA 2961

QY 617 AGACCATCACTTCGAGATCAACAGATCCGACACCGGTCTGTATCAGCTCCGACTCCG 676
DB 2962 AGTGGCTTCGCTTCAGCATCAACAGTGTG-----CGTGGCTTCAGCAAGCGCAAGTACC 3015

QY 677 ACAACATCAACATGTGGATCGGTGATCTTCTATCTTCGCAAGGAGTGTGACCGGTAAAG 736
DB 3016 TCAGCGCTCAGCTTCGCGCTCAACCAAGTGGTCCGCTTCAGCAAGCAACAGTGGCTCAGCC 3075

```





Db 978 CCTGACCCCTGAAGGAGGCGCTGAAGCTGAGCTACCCGACGAGATCAAGGAGATCGAGGG 1037  
QY 702 CTCTATCATCTTCGCCAAGGAGTGGAGGGTAAGGAGATCAACATCTCTGTTCAACTCCTT 761  
Db 1038 CCGTGTGTACTACAAGAACAGCCCATCTACGAGAGCAGCGTGATGACCTATCTAGACGA 1097  
QY 762 GCAGTACACCAACGTCGTCAGGAGTACTGCGGTAAAGCACTGAGATACACAGGAGTA 821  
Db 1098 GAACACCGCCAGGAGGTGACCAAGCACTGAACGACACCAACCGGCAAGTTTCAAGGACGT 1157  
QY 822 CTACATGTCACATCGACTACTTGAACAGATATAGTAAACCACTCCAGACAGATCGT 881  
Db 1158 GAGCCACCTGTAGAGCTGAAGCTCAACCCCAAGATGAAGTGAACATCAAGCTGAGCAT 1217  
QY 882 CTTCACACCAAGCTTACCAACAGACTTCAACAGAGGTTTACAGAGTATCATCAAGCG 941  
Db 1218 CCGTGTACGACACGCGGAGAGCAAGCAAGCATCGGCAAGTGGACCAACCAACAT 1277  
QY 942 TATCAGAGGTAAACCAACGACCAAGAGTCAAGAGTGGTGACATCTCTGTTACTTCGACAT 1001  
Db 1278 CGTAGCGGCGGACACAGCGGCAAGAGCAGTACAGCAGCAACACCCGACGCAACT 1337  
QY 1002 GACTATCAACAGAGGCTACACCTCTTATGATGAAGCAAGACCATGTACGCGCAAA 1061  
Db 1338 GACCCCTGAACCGGCGCGGAGAGAGT-----GAACAAGAACCGCGACTACTACAT 1391  
QY 1062 CCACCTCCAGGAGGATCTACGCGCTCGGTGAGAGTGGTGACATCTCTGTTACTTCGACAT 1121  
Db 1392 CAGCCTGTACATGAGAGCGGCAAGAGCAACCGAGTGCAGATCACTACGCGCGAGAT 1451  
QY 1122 CAACATCATCTTCCAGATCCAGCCCAATGAACAACATTTACTACTACGTTCCCAAGATCTT 1181  
Db 1452 ATACCCCATCATCACCACCAAGACCGTGAACGTTGAACAAGCAACATACAAGCGCTTGGACAT 1511  
QY 1182 CAGTCCCACTCAAGGAGTGAAGATCTCGGTATCTGTTCCATCGGTACCTAC 1236  
Db 1512 CATGCCCAACATCAAGAGCAACCCCATCAGAGCGCTGCACATCAAGACCAAC 1566

## RESULT 15

US-08-471-044-18  
; Sequence 18, Application US/08471044  
; Patent No. 5840868  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Kozel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,044  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pace, Gary M.  
; REGISTRATION NUMBER: 40,403  
; REFERENCE/DOCKET NUMBER: CSC 1695/CIP3/DIV6 - SOLV3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8582  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2004 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..2004  
; OTHER INFORMATION: /note= "Maize optimized DNA  
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"  
US-08-471-044-18  
  
Query Match 6.8%; Score 93.4; DB 2; Length 2004;  
Best Local Similarity 44.3%; Pred. No. 6.le-12;  
Matches 529; Conservative 0; Mismatches 651; Indels 15; Gaps 3;  
  
QY 51 CCGTGTGAAGGACATCATCAAGAGTACTTCAACAACATCAACGACTCCAGATCTCTGTC 110  
Db 378 CCGTGAAGCGCGCATCGGTCCCAAGGCGATCAGCTTCGGCGTGAGCGTGAATACCAACA 437  
QY 111 CCGTGAGACACCGTAAGAACACCTTGGTGCACACCTCCGGTTACACCGCGGAGGTCTCCGA 170  
Db 438 CAGCAGAGCCGTGGGCCCCAGGAGTGGGGCAACGACCGGCAACACAGCAGTTCACAC 437  
QY 171 GGAGGTGACGTCCACAGTGAACCAATCTTCCATTGCGACTTCAAG---CTGGGTTCCTC 227  
Db 498 CGCCAGCGCGGCTACCTGAACCGCAACGCTGCGCTACAACAACGTTGGCCACCGCGCAT 557  
QY 228 CCGTGAGGACAGAGTAAAGTCACTGTCACCCAGACAGAGACATCGTCTACAACTCCAT 287  
Db 558 CTACGACGTGAAGCCCAACCGATCGTTCGTTGAACAACAGCACCACCATCGCCACCATCAC 617  
QY 288 GTACGAGTCTCTCTCCATCTCTTCTGGATCAGATCAACAAGTGGGTCTCCAACTTGCC 347  
Db 618 CGCCAGTCCGATTCACCGCCCTGAACATCAGCCCGCGGAGAGCTACCCCAAGAGGG 677  
QY 348 AGGTACACCATCATCATCGTCTCGTTCAGAACTCCCGTTGGTTCATCGGTATCATCTC 407  
Db 678 CCAGAAGCGCATCGCCCATCAGCAGCATGGAGACTTCAACAGCCACCCCATCCCTGAA 737  
QY 408 CAACCTCTCTGGTCTTCAACCGTGAAGCAGAGGAGTCCGAGCAGTCCCACTTCTC 467  
Db 738 CAAAGAGCAGGTGGAGACACCTGCTGAACAACAAGCCCATGATGCTGGAGACCAACAG 797  
QY 468 CTACGACATCTCCACAACAGCTCTCGTTTACACAAGTGGTTCCTTCCTACCGTACCAA 527  
Db 798 CGACGCGGTCTACAAGATCAAGGACACCCAGGCAAGCATCGTACCGCGCGCGAGTGGAA 857  
QY 528 CAACATGATGGGTACATCAAGATCTACATCAACGCGTAAGCTGATCGACACCATCA-- 584  
Db 858 CCGCGTGTCTCAGCAGATCAAGGCCCAAGACCGCCAGCATCATCTCGTCGACAGCGCGAGG 917

```
Qy 585 ---GGTCAAGAGGATGACCGGTATCAACTCTCCAAAGACCATCACTTCGAGATCAACAA 641
Db ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 918 CGTGGCGGAGAAAGCGGTGGCCGCAAGGACTACGAGAACCCCGAGGACAAAGACCCCGAG 977
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 642 GATCCCGAGACACCGGTCTGATCATCCTCCGACTCCGACACATCAACATGTGGATCCGTGA 701
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 978 CTTGACCTGAAGAGCGCCTGAAGCTGAGCTACCCCGAGAGATCAAGGAGATCGAGGG 1037
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 702 CTTTACATCTTCGCCAAGGAGTTGACGGTTAAGGACATCAACATCTCTGTTTCAACTCCTT 761
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1038 CCTGCTGTACTACAAGAACAAGCCCATCTACGAGACAGCGTGTATCTAGACGA 1097
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 762 GCAGTACACCAAGCTCGTCAAGGACTACTGGGGTACGACCTGAGATACAAAGAGTA 821
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1098 GAACCCGCCAAGAGGTGACCAAGAGTGAACGACACCAACCGGCAAGTTCANGAGCT 1157
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 822 CTACATGTTCAACATCGACTTACTTGAACAGATACATGTACGGCAACTCCAGACAGATCGT 881
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1158 GAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGTGACCATCAAGCTGAGCAT 1217
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 882 CTTCAACACAGAGCTTACAAACAGACTTCAACGAGGTTTCAAGATCATCATCAAGCG 941
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1218 CCTGTACGACAAACCGCGAGAGCAACGACACAGCATCGGCAAGTGGACCAACCAACAT 1277
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 942 TATCAGAGGTAACACCAACGACACAGAGTCAAGAGGTGGTGCATCTCTGTTCTGACAT 1001
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1278 CGTACGCGGGCAACACCGGCAAGAGCAGTACAGCAGCAACACCCCGACGCAACCT 1337
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1002 GACTATCAACAAAGGCTTACAACTGTTTCATGAAGACGAGACCATGTACGCGGACAA 1061
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1338 GACCTGAACACCGACGCGCCGAGGAGAGCT-----GAACAAGAACCGCGACTACTACAT 1391
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1062 CCAGTCCACCGAGGACATCTACGCCATCGGTCTGCTGAGCAGACCAAGGACATCAACGA 1121
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1392 CAGCTGTACATGAGAGCGAGAGACACCCAGTGGGAGATCACCATCGACGGGAGAT 1451
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1122 CAACATCATCTTCAGATCCAGCCATGAACACACTTACTACTAGGTTCCAGATCTT 1181
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1452 ATACCCCATCACCAAGACCGGTGAACGTGAACAAGGACAACTACAAGCGCTGGACAT 1511
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1182 CAAGTCCAACCTCAACGGTGAACATCTCCGGTATCTGTTCCATCGGTACCTAC 1236
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1512 CATCGCCCAACATCAAGAGCAACCCCATCAGAGCCTGCAATCAAGACCAAC 1566
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: November 7, 2002, 13:46:45  
Job time : 74 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 13:24:50 ; Search time 57 Seconds

(without alignments)  
8533.524 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattcagatgaccatcccc.....tctccgagtaataagaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.:

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:
	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1366.2	99.6	1371	10	US-09-350-756-3
2	662	48.3	3876	12	US-10-051-952-4
3	217.4	15.9	1347	10	US-09-350-756-6
4	158.2	14.5	1317	10	US-09-350-756-5
5	168.8	12.3	1338	10	US-09-350-756-1
6	162.4	11.8	1400	10	US-09-350-756-4
7	151.6	11.1	1313	8	US-08-981-087A-6
8	136.6	9.2	702	10	US-09-350-756-10
9	120.4	8.6	1341	10	US-09-350-756-2
10	93.4	6.8	1635	10	US-09-864-761-20241
11	93.4	6.8	1973	10	US-09-864-761-3471
12	92.2	6.7	1397	10	US-09-216-393-343
13	92.2	6.7	1397	10	US-09-216-393-345
14	90.2	6.6	867	10	US-09-216-393-340
15	90.2	6.6	867	10	US-09-216-393-342
16	88.8	6.5	3831	12	US-10-051-952-5
17	82.6	6.0	1959	10	US-09-864-761-4012
18	82.4	6.0	2535	10	US-09-476-242-13
19	81.6	6.0	1075	10	US-09-864-761-19241

c	20	81.6	6.0	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	21	77.6	5.7	2529	10	US-09-476-242-14	Sequence 14, Appl
	22	74	5.4	2298	10	US-09-476-242-22	Sequence 22, Appl
	23	74	5.4	2298	10	US-09-476-242-23	Sequence 23, Appl
	24	74	5.4	2298	10	US-09-476-242-24	Sequence 24, Appl
c	25	73.4	5.4	583	10	US-09-864-761-20772	Sequence 20772, A
	26	71.2	5.2	2523	10	US-09-476-242-15	Sequence 15, Appl
	27	71	5.2	1818	10	US-09-801-368-427	Sequence 427, App
	28	70.8	5.2	2310	10	US-09-476-242-21	Sequence 21, Appl
	29	69.8	5.1	1977	12	US-10-090-624-11	Sequence 11, Appl
	30	69.8	5.1	2541	10	US-09-476-242-12	Sequence 12, Appl
	31	69.4	5.1	2010	12	US-10-032-717-9	Sequence 9, Appl
	32	68.2	5.0	2541	10	US-09-476-242-11	Sequence 11, Appl
	33	67.8	4.9	1518	9	US-09-934-060A-23	Sequence 23, Appl
	34	67.8	4.9	1518	9	US-09-934-060A-29	Sequence 29, Appl
	35	67.8	4.9	1668	9	US-09-934-080A-5	Sequence 5, Appl
	36	67.8	4.9	2159	9	US-09-934-080A-1	Sequence 1, Appl
	37	67.8	4.9	2159	9	US-09-934-080A-3	Sequence 3, Appl
	38	67.6	4.9	1944	10	US-09-864-761-2825	Sequence 2825, Ap
	39	67.4	4.9	2316	10	US-09-476-242-4	Sequence 4, Appl
	40	67.4	4.9	2517	10	US-09-476-242-16	Sequence 16, Appl
	41	67.4	4.9	2517	10	US-09-476-242-17	Sequence 17, Appl
	42	67.4	4.9	2541	10	US-09-476-242-9	Sequence 9, Appl
	43	67.4	4.9	2541	10	US-09-476-242-10	Sequence 10, Appl
	44	66.2	4.8	766	10	US-09-864-761-19608	Sequence 19608, A
	45	66	4.8	2316	10	US-09-476-242-8	Sequence 8, Appl

#### ALIGNMENTS

RESULT 1

US-09-350-756-3

; Sequence 3, Application US/09350756

; Patent No. US20020034521A1

; GENERAL INFORMATION:

; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases

; APPLICANT: John S. Lee

; APPLICANT: Peter Pushko

; APPLICANT: Michael D. Parker

; APPLICANT: Jonathan F. Smith

; APPLICANT: Mark T. Dertzbaugh

; APPLICANT: Leonard Smith

; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine

; FILE REFERENCE: 003/124/SAP RIID 98-21

; CURRENT APPLICATION NUMBER: US/09/350,756

; CURRENT FILING DATE: 1999-07-09

; EARLIER APPLICATION NUMBER: US 60/092,416

; EARLIER FILING DATE: 1998-07-10

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Apple Macintosh Microsoft Word 6.0

; SEQ ID NO 3

; LENGTH: 1371

; TYPE: DNA

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"

US-09-350-756-3

Query Match 99.6%; Score 1366.2; DB 10; Length 1371;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1366; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAATCAGATGACCATCCCATTCACATCTCTCTACACCAACATCCCTGTTGAAG 60

Db 1 GAATCAGATGACCATCCCATTCACATCTCTCTACACCAACATCCCTGTTGAAG 60

Qy 61 GACATCATCAACGAGTACTTCAACAACATCAACGACTCCCAAGATCTGCTCGTGAAG 120

Db 61 GACATCATCAACGAGTACTTCAACAACATCAACGACTCCCAAGATCTGCTCGTGAAG 120

Qy 121 CGTAGAACCTTGGTCGACACTCCGGTTACAAAGCCGAGGTCTCCGAGGGGTGAC 180

Db 121 CGTAGAACCTTGGTCGACACTCCGGTTACAAAGCCGAGGTCTCCGAGGGGTGAC 180

Db 121 CGTAAGAACACCTTGGTGCACACCTCCCGGTTTACAGCGCGAGGTCTCCGAGAGGGTGAAC 180  
QY 181 GTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGA 240  
Db 181 GTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGA 240  
QY 241 GGTAAAGGTATCTGTCACCCAGACAGACATCGTCTACAACTCCATGTACGAGTCCCTC 300  
Db 241 GGTAAAGGTATCTGTCACCCAGACAGACATCGTCTACAACTCCATGTACGAGTCCCTC 300  
QY 301 TCCATCTCTCTTCTGGATCAGATCAACAAAGTGGGTCTCCAACTTCCAGGTTCACCAATC 360  
Db 301 TCCATCTCTCTTCTGGATCAGATCAACAAAGTGGGTCTCCAACTTCCAGGTTCACCAATC 360  
QY 361 ATCGACTCCGTCAAGAACAACTCCGGTGGTCCATCGGTATCATCTCAACTTCCCTGGTC 420  
Db 361 ATCGACTCCGTCAAGAACAACTCCGGTGGTCCATCGGTATCATCTCAACTTCCCTGGTC 420  
QY 421 TTCACCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCC 480  
Db 421 TTCACCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCC 480  
QY 481 AACAAAGCTCTGGTTCACAAAGAGTGGTCTTCGTCAACCGTCCACCAACAACTATGGGT 540  
Db 481 AACAAAGCTCTGGTTCACAAAGAGTGGTCTTCGTCAACCGTCCACCAACAACTATGGGT 540  
QY 541 AACATGAAGATCTACATCAACGAGTGAAGTGCAGACCATCAAGTCAAGGAGTTGACC 600  
Db 541 AACATGAAGATCTACATCAACGAGTGAAGTGCAGACCATCAAGTCAAGGAGTTGACC 600  
QY 601 GGTATCAACTTCTCCAAAGACATCACTTCGAGATCAACAAGATCCCAAGACCGGTCTG 660  
Db 601 GGTATCAACTTCTCCAAAGACATCACTTCGAGATCAACAAGATCCCAAGACCGGTCTG 660  
QY 661 ATCACTCCGACTCCGACACATCAACATGTGGTCCGCTGACTTCTACATCTTCGCCAAG 720  
Db 661 ATCACTCCGACTCCGACACATCAACATGTGGTCCGCTGACTTCTACATCTTCGCCAAG 720  
QY 721 GAGTTGACCGGTGAAGACATCAACATCTTGGTCAACTCTTCGATCAACCAAGTCTGTC 780  
Db 721 GAGTTGACCGGTGAAGACATCAACATCTTGGTCAACTCTTCGATCAACCAAGTCTGTC 780  
QY 781 AAGGACTACTGGGTGAAGACATCAACATCAACAGGAGTACTATGTTCAACATCGAC 840  
Db 781 AAGGACTACTGGGTGAAGACATCAACATCAACAGGAGTACTATGTTCAACATCGAC 840  
QY 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACAGACGTAAC 900  
Db 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACAGACGTAAC 900  
QY 901 AACACGACTTCAACGAGGTACAGATCATCATCAAGCGTATCAGAGGTAAACCAAC 960  
Db 901 AACACGACTTCAACGAGGTACAGATCATCATCAAGCGTATCAGAGGTAAACCAAC 960  
QY 961 GACACAGAGTCAAGGTGTGACATCTTCTGACATGACTATCAACAAACAGGCC 1020  
Db 961 GACACAGAGTCAAGGTGTGACATCTTCTGACATGACTATCAACAAACAGGCC 1020  
QY 1021 TACAACCTGTTATGAAGAACGAGACATGTAGCGGACACCACTCCACGAGGACATC 1080  
Db 1021 TACAACCTGTTATGAAGAACGAGACATGTAGCGGACACCACTCCACGAGGACATC 1080  
QY 1081 TACGCCATCGGTCTGGTGAGCAGCAAGGACATCAACGACAACTATCTTCCAGATC 1140  
Db 1081 TACGCCATCGGTCTGGTGAGCAGCAAGGACATCAACGACAACTATCTTCCAGATC 1140  
QY 1141 CAGCCATGAACAACTTACTACTACGTTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
Db 1141 CAGCCATGAACAACTTACTACTACGTTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
QY 1201 GAGAAATCTCCGGTATCTGTTCCATCGTACTACAGATTCGCTGGTGGTCACTGG 1260  
Db 1201 GAGAAATCTCCGGTATCTGTTCCATCGTACTACAGATTCGCTGGTGGTCACTGG 1260

QY 1261 TACAGACACAACACTACTTGGTTCCTCAACTGTCAAGCAGGTAACCTCCTCTCTGCTGGAG 1320  
Db 1261 TACAGACACAACACTACTTGGTTCCTCAACTGTCAAGCAGGTAACCTCCTCTCTGCTGGAG 1320  
QY 1321 TCCACTTCCACCCACTGGGGATTCGTCCTCCAGTCTCCGAGTATAGGAATTC 1371  
Db 1321 TCCACTTCCACCCACTGGGGATTCGTCCTCCAGTCTCCGAGTATAGGAATTC 1371

## RESULT 2

US-10-051-952-4  
; Sequence 4, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 09/730,237  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: botulinum toxin  
US-10-051-952-4

Query Match 48.3%; Score 662; DB 12; Length 3876;  
Best Local Similarity 68.1%; Pred. No. 5.3e-156;  
Matches 920; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTCTCCCTACCAACAACACTCCCTGTTCAAGGACATCATCAAC 72  
Db 2527 ACAATACCCTTTAATATTTTTCATATATCTAATAATCTTTATTAAGACATATATTAAT 2586  
QY 73 GAGTACTTCAACAACATCAACGACTCCAAGATCTCTCCCTGCAGAACCGTAAGAACACC 132  
Db 2587 GAATATTTCAATAAATATTATGATTCAAAAATTTTGAGCCTACAAAACAGAAAAAATACT 2646  
QY 133 TTGGTCGACACCTCCGGTTACAAACCGGAGGTCTCCGAGGAGGTGAGTCCAGCTGAAC 192  
Db 2647 TTAGTGGATACATCAGGATATATCGAAGTGAAGTGAAGGAGGTGAGTCCAGCTTAAT 2706  
QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCTCCGGTGAGGACAGAGTAAAGTCAATC 252  
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGGTAGTTCCAGGGAGGATAGAGGTAAGATTATA 2766  
QY 253 GTACCCCAAGACGAGACATCGTCTACACATCCATGTACGAGTCTCTCTCCATCTCCTTC 312  
Db 2767 GTAACCCGAATGAATAATTTGTATATATCTATGATGAAGTTTTFAGCATAGTGT 2826  
QY 313 TGGATCAGATCAACAAGTGGGTCTCCAACTTGCAGGTTCACCATCATCGACTCCGCTC 372  
Db 2827 TGGATTAGATTAATAAATGGGTGAAGTATTTACCTGGATATCTATATTAATGATAGTGT 2886  
QY 373 AAGAACAACTCCGGTGGTTCATCGGTATCATCTCCAACTCTCCCTGGTCTTCCACCTGAAG 432  
Db 2887 AAAAAATCACTCAGGTGGAGTATAGTATATTTAGTAAATTTTTTTAGTATTTTACTTTAAA 2946  
QY 433 CAGAAGGAGACTCCGAGCAGTCCATCAACTCTCTCCAGCATCTCCAAACAGCTCCT 492  
Db 2947 CAATATGAGATAGTAGAACAAGTATAAATTTAGTTATGATATATCAAAATAGCTCCT 3006  
QY 493 GGTTAACAAGTGGTTCCTTCCTCCCGTCAACCAACACATGATGGGTAAACATGAAGATC 552  
Db 3007 GGATAGATAAATGTTTTTTTGTAACTGTACTAACAATAATGATGGGAATATGAAGATT 3066  
QY 553 TACATCAACGTTAAGCTGATCGACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTC 612  
Db 3067 TATATAATGGAAATTAATAGATATCTATAAAGTTAAAGAACTAACTGGAATTAATTTT 3126

```
QY 613 TCCAGACCATCACTCTCGAGATCAACAAGATCCCGACACACCGGTCTGATCACTCCGAC 672
    |||||
Db 3127 AGCAAACTATAAATTTGAAATAAATATAATCCAGATACCGGTTGATCTTCAGAT 3186
    |||||
QY 673 TCCGACACATCAACATTTGGATCCGAGCTCTTACATCTTCCGACGAGGTTGGACGGT 732
    |||||
Db 3187 TCTGATACATCAATATGTTGATGAAGAGATTTTATATATTTTCTAAAGAAATTAGATGT 3246
    |||||
QY 733 AAGGACATCAACATCTCTCAACTCTTCAGTACACCAACGCTCTCAAGGACTACTGG 792
    |||||
Db 3247 AAGATATTAATATTAATTAATTAATAGCTTCAATATACATTAATGTTTAAAGATTTATGG 3306
    |||||
QY 793 GGTAAACGCTGATACATCAACAAGGAGTACTATAGTGTCAACATCACTACTTGAACAGA 852
    |||||
Db 3307 GGAATGATTAAGATATAATAAGAAATATATATGTTTAAATATAGATTTATTAATAGA 3366
    |||||
QY 853 TACATGTACGCGCACTCCAGACAGATCGTCTCAACACACAGAGTAAACAACACGACTTC 912
    |||||
Db 3367 TATATGTATGCGAACTCAGCAAAATTTGTTTAAATACAGTAGAATAATAATGACTTC 3426
    |||||
QY 913 AACGAGGTTACAGATCATCAACGCTTACAGAGGTAAACACACACGACGAGTCC 972
    |||||
Db 3427 AATGAAGGATATAAATTAATAAAGAAATCAGAGGAATACAAATGATAGAGTA 3486
    |||||
QY 973 AGAGTGTGATGACATCTCTGACATGACTATCAACAACAAGGCTTACAACTCTTTC 1032
    |||||
Db 3487 CGAGGAGGATATTTATATTTTGTATGACAAATTAATAACAAGCATATAATTTGTTT 3546
    |||||
QY 1033 ATGAAGAAGAGACCATGTACGCGGCAACACCTCCACGAGGAGATCTACGCGCATCGT 1092
    |||||
Db 3547 ATGAGAAATGAACATATGATGCAATTAATCAATAGTACTGAAGATATATATGATAGT 3606
    |||||
QY 1093 CTGCGTGAAGCAGACAGGAGATCAACGCAACACATCATCTTCCAGATCCAGCCAAATGAAC 1152
    |||||
Db 3607 TTAAGAGAACAAACAGGATATAAATGATTAATATATATTTCAATATCAACCAATGAAT 3666
    |||||
QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGCTGAGAGACATCTCC 1212
    |||||
Db 3667 AATACTTATTTATGAGTCTCAAAATTTTAAATCAAAATTTTATGAGAGAAATATTTCT 3726
    |||||
QY 1213 GGTATCTGTCCTCGGTACTACAGATTCGCTGCTGGGTGAGTCTGAGTGGTACAGACAAAC 1272
    |||||
Db 3727 GGAATGTTTCAATAGTACTATCTGTTAGTCTGGAGGTGATTTGGTATAGACAAAT 3786
    |||||
QY 1273 TACTTGGTTCACATGTCAGACGAGGTACTACGCTCTTGTGAGGAGTCCACTTCCACC 1332
    |||||
Db 3787 TATTTGGTCTACTGTGAAGCAAGGAAATATGCTTCAATATTAATAGAAATCAACATCAACT 3846
    |||||
QY 1333 CACTGGGATTCGTCGCGGATCTCCGAGTAA 1362
    |||||
Db 3847 CATTTGGGTTTTGTACCTGTAAAGTGAATAA 3876
    |||||
```

## RESULT 3

```
US-09-350-756-6
; Sequence 6, Application US/09350755
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
```

```
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 6
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-6
```

```
Query Match 15.9%; Score 217.4; DB 10; Length 1347;
Best Local Similarity 52.5%; Pred. No. 4e-45;
Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;

QY 43 AACAACTCCCTGTTGAGGACATCATCAGAGTACTTCAACAACATCAACGACTTCCAAG 102
    |||||
Db 4 AAGACACCACTCCCTGATCCAGGCTCTCAACAACATCTCTCCAACTCTCCCAACGCC 63
    |||||
QY 103 ATCTGTCCCTGCGAAGACCCGTAAGAACACCTTGGTGCACACCTCCCGTTACACGCCGAG 162
    |||||
Db 64 ATCTGTCCCTGCTCTACCTACCGTGGTGTCTGATCGACTCTCTCCGTTACGGAGCCACC 123
    |||||
QY 163 GTCTCCGAGGAGGTTGACGTCCAGCTGAACCCAAATCTTCCCATTCGACTTCAAGTGGT 222
    |||||
Db 124 ATCAACCTCGGTTCCGACGTCATCTTCAAGGACATCGGTAAACGTCAGTTCAGTGTGAAC 183
    |||||
QY 223 TCCCTCCGTTGAGGACAGAGGTAAAGGTCTCGTCCACCCAGAGAGAACATCGTCTACAAAC 282
    |||||
Db 184 AACTCC-----GAGAACTCCAACATCACCGCCACCAGTCCCAAGTTCGTCGTACGAC 237
    |||||
QY 283 TCCATGACAGAGTCTTCTGATCTCTGATCAGATCAGAAATCAACAAGTGGTCTTCCCAAC 342
    |||||
Db 238 TCCATGTTCAACAACACTTCTCCATCAACTTCTGGTCCGTAACCCCAAGATACAAACAAAC 297
    |||||
QY 343 -----TTGCCAGGTTACACCATCATCGACTTCCGTCGTAAGAACAACTCC 384
    |||||
Db 298 GACATCCAGACCTACCTGCGAGACGAGTACACCATCATCTCTCTGTATCAAGACGACTCC 357
    |||||
QY 385 GGTGTGTCATCGGTATCTCTCAACTTCTGTTCTTCCCTTCAACCTGAAGACAGACGAGGAC 444
    |||||
Db 358 GGTGTGAGAGTCTCTCAATCAAGGAAACCGTATCATCTGGACCTGATGACCTCAACGCC 417
    |||||
QY 445 TCCGAGGACGTCATCAACTCTCTCTAGGACATCTCAACAACGCTCTCTGTTA--CAAC 501
    |||||
Db 418 AAGTCCAGTCCATCTCTCTGAGTACTCTCAAGGACACATCTCCGACTACATCAAC 477
    |||||
QY 502 AAGTGGTCTTCTGTCACCGTCAACCAACATGATGGGTAAATGAGATCTTACATCAAC 561
    |||||
Db 478 AAGTGGTCTCTCATCAACATCAACCAACGCGTCTGGTAACGCCAACATCTACATCAAC 537
    |||||
QY 562 GGTAAAGTGTGACACACCATCAAGTCAAGGAGTTCACCGGTATCAACTTCTCCAAAGACC 621
    |||||
Db 538 GGTTCCTGAAAGAGTCCGAGAGATCTCTGAACCTGACCGTATCAACTCTCTCCAAACGAC 597
    |||||
QY 622 ATCACTTCCGAGATCAACAAGATCCCAAGACCGGTGTGATCACTCCGACTCCCGACAC 681
    |||||
Db 598 ATCGACTTCAAG-----CTGATGAACGTGTACCGACACACACC 633
    |||||
QY 682 ATCAACATGTGGATCCGTAAGTCTTACATCTTCCCAAGAGGTTGGCGGTAAAGGACATC 741
    |||||
Db 634 AAGTGTGCTGATCAAGGACTTCAACATCTTGGTGTGAGCTGAACGCCACCGAGGTC 693
    |||||
QY 742 AACATCTGTTCAACTCTCTGAGTACACCAACGTCGTCAAGGACTACTGGGTAAACGAC 801
    |||||
Db 694 TCCCTCCCTGTACTGTGATCCAGTCTCCCAACACACCTGAAAGGACTTCTGGGGAACCCA 753
    |||||
QY 802 CTGAGATCAACAAGGAGTACTATAGTCAACATCGACTACTTGAACAGATACATGTAC 861
    |||||
Db 754 CTGGGTTACGACACCCAGTACTACTCTGTTCAACCAAGGATGATGAGAAACATCTACATCAAG 813
    |||||
QY 862 GCCAACTCCA---GACAGATGCTCTTCAACACCAAGAGCTAAACAACAACTTCAACAGAG 918
    |||||
Db 814 TACTTCTCCAGGCTCTCCATGGTGGTGAACGCCCTCGTACCAACTTCAACAACGCCGCC 973
    |||||
```

```
QY 919 GGTACAGATCATCATCAAGCGTATACAGAGTAAACACACAGACACACAGAGTACAGAGT 978
Db 874 ATCACTACAGAACCTGTACCTGGTCTGGTCTGGTCTGGTCTGGTCTGGTCTGGTCTGGT 933
QY 979 GGTACAGATCATCATCAAGCGTATACAGAGTAAACACACAGACACACAGAGTACAGAGT 1038
Db 934 CGTACATCAACAGACACACATCTCCGCTGAGGCTGACTACATCTACCTGACATCGAC 993
QY 1039 AAGGAGACATCTAGCGCGACACACATCTCCAGGAGGACATCTACGCGATCGGCTCGGT 1098
Db 994 AACATCTCGGAGGCTGTACCGTGTACCTGTACCTGTACCTGTACCTGTACCTGTACCTGT 1053
QY 1099 GACACACACAGACATCAAGACACACATCTCTCCAGATCCAGCCCAACACACACACT 1158
Db 1054 CAGCTTCTCTGGCCCAATCAACAGACACACCTACCTTCTACGAGCGCTGACATCAAG 1113
QY 1159 TACTACTAGGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAAGACATCTCCGGTATC 1218
Db 1114 AAGTACTAGGAGAGACACCTTACAACTGTGAGATCTGTGCGAAGGACACCAAGACC 1173
QY 1219 T 1219
Db 1174 T 1174

RESULT 4
US-09-350-756-5
; Sequence 5, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 5
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-5

Query Match 14.5%; Score 198.2; DB 10; Length 1317;
Best Local Similarity 56.6%; Pred. No. 2.5e-40;
Matches 496; Conservative 0; Mismatches 333; Indels 48; Gaps 5;

QY 27 CATCTTCTCTACACCAACTCCCTGTTGAAGGACATCATCAAGAGTACTTCAACAA 86
Db 6 CAGGATGTCCTACACCAAGCAGACATCTGATCTTGTACTTCAACAAAGCTGTACAA 65
QY 87 CATCAAGCATCAAGATCTCTGCTGAGACCGGTAAAGACACTTGTGCGACACTC 146
Db 66 GATCAGGACAACTCCATCTTGGACATGAGATACGAAACAAATAGTTCATCGACATCTC 125
QY 147 CGGTACACCGCGAGGTCTCGGAGAGGGTCACTCCAGTCAACCAATCTTCCCAT 206
Db 126 CGGTACCGGTCAACATCTCCATCAACGGTCACTTACTCTACATCTCCACCAATAGAA 185
QY 207 CGACTTCAAGCTGGTCTCTCGGTGAGGACAGAGTAAAGTCACTGTCACCCAGACGA 266
Db 186 CCAGTTCGGATCTACTCTCTCC-----AAGCCTTCGAGGTCAACATCGCTCAGACAA 239
```

```
QY 267 GAACATCTCTACAACTCCATGTACGAGTCTCTCCATCTCTCTCTCTCTCTCTCTCTCTCT 326
Db 240 CGACATCTCTACAACTCCATGTACGAGTCTCTCCATCTCTCTCTCTCTCTCTCTCTCTCT 299
QY 327 CAAGTGGGTCTCCAACTT-----GCCAGGTTACACCATCATCGACCTCCGTC---AA 374
Db 300 AAGTACTTCAACAAAGTCAACCTGATAGGATACACCATCATCGACCTCCGTC---AA 359
QY 375 GAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 434
Db 360 CAATCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 419
QY 435 GAACGAGGACTCCGAGCAGTCCATCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 494
Db 420 CACCCCGGTCAACATCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 479
QY 495 TTA---CAACAGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 551
Db 480 CTACATCAACAAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 539
QY 552 CTACATCAACGTTAGCTGATCGACACCATCAAGTCAAGGTTGACCGGTATCACTT 611
Db 540 CTACATCAACGTTAGCTGATCGACAGAGTCCATCTCTCAACTTGGGTGACATCTCCACGT 599
QY 612 CTCCAGGACCATCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 671
Db 600 CTCCGACAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 646
QY 672 CTCCGACAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 731
Db 647 -----ACGTCTGGGATCCGTTACTTCAAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 692
QY 732 TAAGGACATCAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 791
Db 693 GACCGAGATCGAGACCTTGTACTCCACGAGCTTGACCCATCCATCTCTCTCTCTCTCTCTCTCT 752
QY 792 GGGTAAAGCTGAGATCAACAGGAGTACTATAGTGTACATCGTCACTCTCTCTCTCTCTCTCTCT 851
Db 753 GGGTAAAGCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 812
QY 852 ATACATGTACGCAACTCCACAGATCGTCTCTCAAC 888
Db 813 GTCCATCAACCGAGACTCCAACTCTTGAACATCAAC 849

RESULT 5
US-09-350-756-1
; Sequence 1, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 1
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-1
```

```
Query Match      12.3%; Score 168.8; DB 10; Length 1338;
Best Local Similarity 54.8%; Pred. No. 5.6e-33;
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;

Qy 64 ATCATCAACGAGTACTTCAACACATCAACGAGTCCCAAGATCCTGCTCCGAGAACCGT 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 ACCTTCACTGAATACATCAAGACATCAATACCTCCATCCTGAACCTGGCTACGAA 86

Qy 124 AAGAACACCTTGGTGACACCTCCGCTTACAGCGCGAGGTCTCCGAGGAGGTGACGTC 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87 TCCAACTACCTGATCGACCTGCTCGCTAGCTTCCAAATCAACATCGGTTCTAAAGTT 146

Qy 184 CAGCTGAACCAATCTCCATTCGACTCAAGCTGGGTCTCCGCTGAGGACAGAGGT 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 147 AACTTCGATCCGATCGNCAAGATCAGATCCAGCTG-----TTCAATCTGGAATCTCC 200

Qy 244 AAGGTATCCTACCCAGACAGAGACATCGTCTACAACTCCATCCATCCTAGCGCTTCCTCC 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 AAAATCGAAGTTATCTCTGAAGATGCTATCGTATACAACTCTATGTACGAAAACCTCTCC 260

Qy 304 ATCTCCCTCTGGATCAGATCAACAAGTGGGTCTCCAACTGCC-----AGGTTAC 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 ACCTCCCTCTGGATCCGTATCCGGAATACTTCAACTCCATCTCTGAACTGAAATAC 320

Qy 355 ACCATATCAGCTCCGCTCAAGAACTCCGGTTGGTCTCCATCGGTATCATCTCCAACTTC 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 ACCATATCAACTGCTATGGAACAAATCTCTGGTTGGAAGTATCTCTGAACTACGGTGAA 380

Qy 415 CTGGTCTTCACCTGAGCAGACGAGGACTCCGAGCAGTCCATCACTTCTTCCATAGAC 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 ATCATCTGGACTCTGAGGACATCAGGAATCAACACGCGTGTGATTAACAACTACTCT 440

Qy 475 ATCTCCCAACACGCTCTCTGTTTAAACAA-----GTGGTTCTTGGTCACTCCACCAACAC 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTCTGTACCATCAACCAAT 500

Qy 532 ATGATGGGTAAACATGAGATCTACATCAACGGTAACTGATCGACACCAATCAAGTCAAG 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 CGTCTGAATCACTCAAAATCTACATCAACGGCGCTGTGAGCAGCAAAACCGATCTCC 560

Qy 592 GAGTTGACCGGTATCACTCTCCAGACATCACTCCCTGAGATCAACAGATCCAGAC 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 AATCTGGGTACATCCAGCTTCTATACATCATGTTCAACTGGACGGTTGCTGTGAC 620

Qy 552 ACCGGTCTGATCACTCCGACTCCGACACATCAACATGATGATCGGTGATCTTACATC 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 621 ACTC-----ACCGCTACATCTGGATCAAAATCACTCAATCTG 656

Qy 712 TTCGCCAAGGAGTGGAGGTAAGGACATCAACATCCTCTTCAACTCCTTGCAGTACACC 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 TTCGACAAAGAACTGAACGAAAGAAATCAAGACCTGTACGACAAACGATCCATCT 716

Qy 772 AACGTCGTCAAGGACTACTGGGTAACGACCTGAGATACAAAGAGGAGTACATGCTC 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 GGTATCCTGAAGACTTCTGGGTGACTACCTCGAGTACGACAAACCGTACTACATGCTG 776

Qy 832 AACATCGACTACTGAACATACATGATAGCGCAAC 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 AATCTGTACGTCGGAACAAATACGTTGAGCTCAAC 812

RESULT 6
US-09-350-756-4
; Sequence 4; Application US/09350756
; Patent No. US2002034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith

; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP R1ID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; EARLIER FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 4
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-4

Query Match      11.8%; Score 162.4; DB 10; Length 1400;
Best Local Similarity 54.4%; Pred. No. 2.3e-31;
Matches 468; Conservative 0; Mismatches 341; Indels 51; Gaps 5;

Qy 16 ATCCCATTCACATCTTCTCTACACCAACAACCTCCCTGTTCAAGGACATCATCAACGAG 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ATTCCITTTAAGCTTTCTTCTTATACAGATGATAAAATTTTAATTTCTTCTTCAACAG 126

Qy 76 TACTTCAACAAACATCAACGACTCCAAAGATCCTGCTCCGTCAGAACCGTAAGAACACCTTG 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 TTCTTCAAGAGAAATTAAGTCTTCTCCGTTTAAACATGAGATCAAGAATGATAAATAC 186

Qy 136 GTGACACCTCCGGTTACAAACGCGAGGCTCCGAGGAGGTTGAGTCCAGTGAACCCA 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 GTGACACCTCCGGTTACGACTCCAAATATCAACATTAACGGTGGAGCTGTACAAAGTACCCA 246

Qy 196 ATCTTCCATTCGACTTCAAGTGGGTCTCTCCGGTGGAGGACAGAGGTAAGTCACTGCTC 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 ACTAACAACAAACCAATTC-----GGTATCTACAAGACAAGCTTACTGAGCTGAACATC 300

Qy 256 ACCGACAGAACGAAACATGCTCTACAACTCCATGTAGGAGTCTTCTTCCATCTCTCTGG 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TCTCAAAACGACTACATTTATCTACGACAAACAGTACAAAGAACTTCTCTATTTCTTCTGG 360

Qy 316 ATCAGAAAT-----CAACAAGTGGGTCTCCAACTTGCCAGGTTACACCACTATC 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTCAGGATTCCTAACTACGACACAGATCGTCARAGTTAACAACGAGTACACTATCATC 420

Qy 364 GACTCCGCTCA---AGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCTGGTC 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 AACTGTATGAGACAAACAACCTCCGGTTGGAGGTCTCTCTTAACCAACAACGAGATCAT 480

Qy 421 TTCACCTGGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTTACGACATCTCC 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 TGGACCTTGAACACACGAGGAGGTATTAACCAAAAGTTAGCATTAACCTACGCTTAACGCA 540

Qy 481 AACACGCTCCTCTGTTA---CAACAAGTGGTTCTTGGTCAACCGTCAACCAACAACATGATG 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 AACGGTATTTCTGACTACATCAACAAGTGGATTTTGGTCACTATCACTAACGACAGATTA 600

Qy 538 GGTACATGAAGATCTACATCAACGCTAGCTGACGACACCATCAAGTCAAGGAGTTG 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 GGTGACTCTAAGCTTTACATTAACGGTAACCTTAATGACCAAAAGTCCATTTTAACATTA 660

Qy 598 ACCGGTATCAACTTCTCCAAAGACCATCACTCCGAGATCAACAAGATCCGACACACCGGT 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 GGTAACATTCACGTTCTGACACATCTTATTCAGATCGTTTAACGAGTTACACGAGA 720

Qy 658 CTGATCACTCCGACTCCGACCAACATCAACATGTGGATCCGCTGATCTTCTACATCTTCGCC 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 T-----ACATTTGGCATTTAGATTACTTCAACATTTTCGAC 753

Qy 718 AAGGAGTTGACGCTAAGGACATCAACATCTCTGTTCACTCTTCACTCTGAGTACACCAACGTC 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754 AAGGAGTTAGACGAGACCGAGATTCAACTTTATACGACGACGAACTTAACACCACTATTT 813

Qy 778 GTCAAGGACTACTGGGTAACGACCTTGAGATCAACAAGGAGTACTACATGTTGTTCAACATC 837
```



Db 814 TTGAAGGACTTCTGGGTAACACTACTTGCTTTTACGACAGGAATACTACTATTATAACGTTG 873

QY 838 GACTACTTGAACAGATACAT 857

Db 874 TTAAGCCAAACAACATTCAT 893

## RESULT 7

US-08-981-087A-6  
; Sequence 6, Application US/08981087A  
; Patent No. US20020081304A1  
; GENERAL INFORMATION:  
; APPLICANT: Elmore, Michael J.  
; APPLICANT: Nauchline, Margaret L.  
; APPLICANT: Minton, Nigel P.  
; APPLICANT: Pasechnik, Vladimir A.  
; APPLICANT: Titball, Richard W.  
; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 No. US20020081304A1th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981.087A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01409  
; FILING DATE: 12-JUN-1996

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9511909.5  
; FILING DATE: 12-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-688  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-981-087A-6

Query Match 11.1%; Score 151.6; DB 8; Length 1313;  
Best Local Similarity 53.3%; Fred. No. 1.1e-28;  
Matches 465; Conservative 0; Mismatches 359; Indels 48; Gaps 5;  
QY 32 TCTCTACACCAACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAACATCA 91  
Db 10 TGTCTTACACTAGGACAAAATCTTGATCTTCAACAACATCTGACAAAATCA 69  
QY 92 ACAGCTCCAGATCCCTGTCGAGACCGTGAAGAACACCTTGGTCGACACCTCCGGTT 151  
Db 70 AAGACAACCTTATCTCTGGACATCGCTAGCAAAACAACAAATTCATCGACATCTCTGGCT 129  
QY 152 ACAACGCGGAGGTCCTCGAGGAGGTGAGCTCCAGCTGAACCAATCTCCATTCGACT 211  
Db 130 ATGGTTCTACATCTATACACGGTGACGCTGACTACTCTA-----CTAACCGCA 183

QY 212 TCAGCTGGGTTCTCCGGTGAGGACAGAGGTAGGTGTCATCGTCACCCAGAACGAGACA 271  
Db 184 ACCAGTTCCGTTATCTCTTCTAAACCGTCTGAAGTAAACATCGCTCAGAACAGACA 243  
QY 272 TCGTCTACAACTCCATGTACGAGTCCCTTCCATCTCTTCTGTGATCAGAAATCAACAAGT 331  
Db 244 TCACTACACGGTCGTACGAGAACTTCTATCTCTTCTGGTTCGTATCCGAAAT 303  
QY 332 GGGTCTCCAACTT-----GCCAGGTACACCATCATCGACTCCGGT----AAGAACA 379  
Db 304 ACTTCACAAAAGTTAAAGTGAACGAAACAAATACATATCATCGACTGATCCGTAACAACA 363  
QY 380 ACTCGGTTGGTCGATCGGTATCATCTCCAACTTCCGGTCTTCCACCTGAACGAGAACG 439  
Db 364 ACTCTGGTTGGAAATCTCTCTGAACTACACAAATCATCTGGACTCTCGAGGACATG 423  
QY 440 AGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCCAAACAGGTCCTCGGTAC- 498  
Db 424 CTGGTAACAACGAGAACTGGTTTCAACTACACTCAGATGATCTCTATCTGAACTACA 483  
QY 499 --AACAAGTGGTTCTGTGTCACCGTCACCAACACATGATGGGTACATGAAGATCTACA 556  
Db 484 TTAATAAATGGATCTTGGTTACTTACTTACTTAAACCCCTCTGGGTAATCTCTGATCTACA 543  
QY 557 TCACGGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCA 616  
Db 544 TCAACGGTAACCTGATGATGAAAAATCTATCTTAACCTGGGTGACATCCAGCTTCTG 603  
QY 617 AGACCATCACTTCGAGATCAACAGATCCAGACACCGGTCTGATCACCCTCGACTCCG 676  
Db 604 ACAACATCCTGTTCAAAATCTTGGTTGCAACGACACCGGTT----- 645  
QY 677 ACAACATCAACATGTGGATCGGTGACTTCTACATCTTCGCAAGGAGTTGGACGGTAAG 736  
Db 646 -----ACGTTGGTATCGTTACTTCAAAGTTTTCGACACTGAACTGGTAAACTG 696  
QY 737 ACATCAACATCCTGTTCACTCTTGCAGTACACCAACGCTGTCAGGAGTACTGGGTA 796  
Db 697 AAATCGAAACTCTGTACTCTCAGCAACCGGACCGCTATCTCCTGAAAGACTTCTGGGTA 756  
QY 797 AGGACTGAGATACAAACAGGAGTACTACATGTCGCAACATCGACTTCTGAACAGATACA 856  
Db 757 ACTACCTGCTGTACAAACAAACGTTACTTACCTGCTGAACCTGCTCCGGACTGACAAATCTA 816  
QY 857 TGTAGCCAACTCCAGACAGATCGTCTTCAAC 888  
Db 817 TCACCTCAGAACTCTAACTTCTGTAACATCAAC 848

## RESULT 8

US-09-350-756-10  
; Sequence 10, Application US/09350756  
; Patent No. US20020034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350.756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092.416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 10  
; LENGTH: 702  
; TYPE: DNA

```
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-10

Query Match      9.2%; Score 126.6; DB 10; Length 702;
Best Local Similarity 53.6%; Pred. No. 1.5e-22;
Matches 382; Conservative 0; Mismatches 289; Indels 42; Gaps 4;

QY 54 ATCATCAGAGTACTTCACACATCAACGACTCCAGACTCCTGCTCCGTCGAGACCGT 123
DB 19 ACCCTCAGTAACTACACAGACATCAATACCTCCAGACTCCGCTGAGCTGCGCTACGAA 78

QY 124 AAGAACACCTTGGTCGACACCTCGGTTACACGCGGAGGTCTCCGAGAGGGTGCAGTTC 183
DB 79 TCCAAATCAGCTGATCGACCTGCTCGCTACGCTTCCAAATCAACATCAATCTTAAAGTT 138

QY 184 CAGCTGAACCAATCTTCCATTCGACTTCAAGCTCGGTTCCCGGTGAGGACAGAGT 243
DB 139 AACTTCGATCGATCGACGACAGAAATCAGATCCAGCTG-----TTCAAATCTGGAATCTTCC 192

QY 244 AAGGTCTATCGTCAACCCAGAACGATCGTCTACAACCTCCATGTACGAGTCCCTTCTCC 303
DB 193 AAATCGAAGTTATCCGTAAGAAATGCTATCGTATACAACCTCTATGTACGAAACATCTCC 252

QY 304 ATCTCCCTTCTGGATCAGAAATCAACAGTGGTCTCCAACTTGGC-----AGGTTCAC 354
DB 253 ACCTCTCTTCTGGATCCGATATCCCGAAATCTTCAACTCCATCTCTCTGAACATGAATAC 312

QY 355 ACCATCATCGACTCCGTCAGAACAACTCCGGTGTGGTCCATCGGTATCTCCAACTTC 414
DB 313 ACCATCATCAACTCGATCGAAACAAATCTCTGGTGGAAAGTATCTCTGAACATCGGTGAA 372

QY 415 CTGGTCTTACCCCTGAGCAGACGAGGACTCCGAGGAGTCCATCAACTCTCCCTACGAC 474
DB 373 ATCATCTGGACTCTGCGAGGACACTCAGGAAATCAACAGCGGTCTTGTATTAATAACTCT 432

QY 475 ATCTCCAAACAGCTCTCTGGTTTACAAACAA---GTGGTCTTCTGTCACCGTFCACCAACAC 531
DB 433 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTCTGCTTACCATCAACACAT 492

QY 532 ATGATGGGTACATGAGATCTACATCAACGGTAAAGTGTATGACACACCATCAAGGTCAAG 591
DB 493 CGTCTGAATACTCAAAATCTACATCAACGGCGGCTGTGATGACGAGAAACCGATCTCC 552

QY 592 GAGTGTACCGGTATCAACTCTCCAGACCATCACTTCGAGATCAACAGATCCCGAGAC 651
DB 553 AATCTGGGTACATCCAGCTCTTATATACATCATGTTCAAACTGGACGGTGTCTGTGAC 612

QY 652 ACCGGTCTGATCACTCCGATCCGATCCGACATCAACATGTGGATCGGTGACTTCTACATC 711
DB 613 ACTC-----ACCGCTACATCTGGATCAAAATACATCTCAATCTG 648

QY 712 TTCGCCAAGGAGTTGGAGGTAAAGGACATCAACATCTCTGTTCAACTCCCTTGCA 764
DB 649 TTCGCAAGAAGACTGAACGAAAGAAATCAAGACCTGTACGACCAACACAGTA 701

RESULT 9
US-09-350-756-2
; Sequence 2, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09

; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 2
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2

Query Match      8.8%; Score 120.4; DB 10; Length 1341;
Best Local Similarity 50.7%; Pred. No. 6.9e-21;
Matches 412; Conservative 0; Mismatches 346; Indels 54; Gaps 3;

QY 54 ATCATCAACGAGTACTTCACACATCAACGACTCCAGACTCCTGCTCCGTCGAGACCGT 123
DB 10 ATGCCCAACAAATACAAATCCGAAATCCCTGAACAATATCATCTCTGAACCTGGTGTACAAA 69

QY 124 AAGAACACCTTGGTCGACACCTCGGTTACACGCGGAGGTCTCCGAGGAGGTGCAGTTC 183
DB 70 GACAACAATCTGATCGATCTGCTGTTACGGTGTCTAAAGTTGAGTATACGACGCTGTT 129

QY 184 CAGCTGAACCAATCTTCCATTCGACTTCAAGCTGGTTCCTCCGCTGAGGACAGAGT 243
DB 130 GAAGTGAATGACAGAACCACTTCAAACTGA-----CCTCTTCCGCTAACTCT 177

QY 244 AAGGTCTATCGTCAACCCAGAACGATCGTCTACAACCTCCATGTACGAGTCCCTTCTCC 303
DB 178 AAGATCGGTGTTACTCAGATCAGAACATCATCTTCAACTCCGTTATTCCTGGACATCTCT 237

QY 304 ATCTCCCTTCTGGATCAGAAATCAACAGTG-----GGTCTCCAACTTG 345
DB 238 GTTTCCTTCTGGATCTGATCCGAAATACAGAACGAGCGGTATCCAGAAATACATCCAC 297

QY 346 CCAGGTACACCATCATCTGCTCCGTCAGAACAACTCCGGTGTGGTCCCTCGTATCATC 405
DB 298 AATGAATACACCATCATCACTGATGAGAAATCACTCTGTTGGAGATCTCCATCCGC 357

QY 406 TCCAACTCTCTGGTCTTCAACCTGAGCAGACGAGAGATCCGAGCAGTCCATCAACTTC 465
DB 358 GGTACCGGTATCATCTGGACTCTGATCATCAACGGTAAAGACCAAACTCTGTATTCTTC 417

QY 466 TCCTACGACATCTCCAAACAGCTCCCTGGTTTACAACAGTGGTCTTCGTCACCGTCCAC 525
DB 418 GAATACACATCTCCGTAAGACATCTCTGAATACATCAATCGCTGGTCTTCGTTACCATC 477

QY 526 AACAACTGATGGGTAAACATCAAGATCTACATCAACGGTAAAGCTGATGACACCATCAAG 585
DB 478 ACCAATAACCTGAAACAATGCTAAATCTACATCAACGGTAAACTGGAATCTAATACCGAC 537

QY 586 GTCAAGGAGTGTACCGGTATCAACTCTTCCAAACCATCACTCCCTTCGAGATCAACAGATC 645
DB 538 ATCAAGACATCTCGTGAAGTATCGGTAAAGTATCGGTAAAGTATCATCTTCAAACTGGACGTGAC 597

QY 646 CCAGACACCGGTCTGATCACTCCGACTCCGACATCAACATCAACATGTGGATCGGTGACTTC 705
DB 598 ATCGATCGTACCCAGTTTATC-----TGGATGAATACTTC 633

QY 706 TACATCTTCCGCAAGGAGTGGAGGTAAAGGACATCAACATCTGTGTTCACCTCTTGGAG 765
DB 634 TCCATCTTCAACACCGCAACTCTCTGATCCCAATATCGAAGACGGTACAAAGATCCAGTCT 693

QY 766 TACACCAACGTCGTCGAGGACTACTGGGTGAAGGACTGAGATACAAACAGGAGTACTAC 825
DB 694 TACTCCGAATACCTGAAAGACTCTCTGGGTAAATCCCGCTGATGTACAAACAAAGATACTAT 753

QY 826 ATGTCAACATCGACTACTTGAACAGATACAT 857
DB 754 ATGTTCAATGCTGTGTAACAAGAACTCTTTACAT 785
```





RESULT 14

```

US-09-216-393-340
; Sequence 340, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 340
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
US-09-216-393-340

Query Match
Best Local Similarity 51.1%; Score 90.2; DB 10; Length 867;
Matches 212; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 734 AGGACATCAACATCCTGTTCAACTCCTTGCAGTACACCAACGCTGCTCAAGGACTACTGGG 793
Db 451 AGTTATCCCAAGTACGAGTACAAACGACGACATCAACTACCACTACTACCACTACC 510
QY 794 GTAACGACCTGAGATACAAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGAT 853
Db 511 ACCACTAGGACTACTACTACAACTACGAGACCAACCACTACCAACCAACCAACCA 570
QY 854 ACATGTAGCCCACTCCAGACAGATCGTCTTCAACAGCAGCACTACCAACCAACCA 913
Db 571 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 630
QY 914 ACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGGTAAACCAACCAACCAACCA 973
Db 631 ACAACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 690
QY 974 GAGTGGTGACATCCTGTACTTGCAGATGACTATCAACAAAGGCTTCAACCTGTTCA 1033
Db 691 ACCACTAGGACTTACCAACAACCTACTACTACCAACTACCAACCAACCAACCAACCA 750
QY 1034 TGAAGACGAGACCATGTAGCCGACCAACCACTCCACCGAGGACATCTAGCCATCGGTC 1093
Db 751 ACAACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 810
QY 1094 TCGGTGACGACCAAGGACATCAACGACCAACCAATCATCTTCCAGATCCAGCCAA 1148
Db 811 ACTACAGGACTACGACACCATCTACGACGACATCCACCACCACTACCCCTCGATT 865

RESULT 15
US-09-216-393-342
; Sequence 342, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 342
; LENGTH: 867

```

```

; TYPE: DNA
; ORGANISM: Toxoplasma gondii
US-09-216-393-342

Query Match
Best Local Similarity 51.1%; Score 90.2; DB 10; Length 867;
Matches 212; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 734 AGGACATCAACATCCTGTTCAACTCCTTGCAGTACACCAACGCTGCTCAAGGACTACTGGG 793
Db 451 AGTTATCCCAAGTACGAGTACAAACGACGACATCAACTACCACTACTACCACTACC 510
QY 794 GTAACGACCTGAGATACAAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGAT 853
Db 511 ACCACTAGGACTACTACTACAACTACGAGACCAACCACTACCAACCAACCAACCA 570
QY 854 ACATGTAGCCCACTCCAGACAGATCGTCTTCAACAGCAGCACTACCAACCAACCA 913
Db 571 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 630
QY 914 ACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGGTAAACCAACCAACCAACCA 973
Db 631 ACAACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 690
QY 974 GAGTGGTGACATCCTGTACTTGCAGATGACTATCAACAAAGGCTTCAACCTGTTCA 1033
Db 691 ACCACTAGGACTTACCAACAACCTACTACTACCAACTACCAACCAACCAACCAACCA 750
QY 1034 TGAAGACGAGACCATGTAGCCGACCAACCACTCCACCGAGGACATCTAGCCATCGGTC 1093
Db 751 ACAACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 810
QY 1094 TCGGTGACGACCAAGGACATCAACGACCAACCAATCATCTTCCAGATCCAGCCAA 1148
Db 811 ACTACAGGACTACGACACCATCTACGACGACATCCACCACCACTACCCCTCGATT 865

Search completed: November 7, 2002, 14:41:50
Job time : 68 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:21:30 ; Search time 3155 Seconds  
(without alignments)  
10925.626 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattcacgagaccatccc.....tctccgagtaataaggaaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main.\*  
1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1371	100.0	1371	23	US-09-611-419A-9
2	1371	100.0	1371	34	US-09-910-186A-9
3	1366.2	99.6	1371	17	US-09-350-756-3
4	685.2	48.5	3876	1	PCT-US97-15394-59
5	685.2	48.5	3876	11	US-08-704-159-59
6	685.2	48.5	3876	42	US-10-205-516-5
7	665.2	48.5	3876	43	US-10-271-012-59
8	665.2	48.5	3950	13	US-08-954-302-2
9	663.2	48.4	3906	42	US-10-205-516-19
10	662	48.3	3876	29	US-09-730-237-4
11	662	48.3	3876	39	US-10-051-952-4
12	649.8	47.4	1502	1	PCT-US97-15394-61
13	649.8	47.4	1502	11	US-08-704-159-61
14	649.8	47.4	1502	43	US-10-271-012-61
15	420.2	30.6	1374	23	US-09-611-419A-11
16	417	30.4	1374	34	US-09-910-186A-11
17	217.4	15.9	1347	17	US-09-350-756-5
18	217.4	15.9	1368	23	US-09-611-419A-17
19	217.4	15.9	1368	34	US-09-910-186A-17
20	199.8	14.6	1317	23	US-09-611-419A-15
21	199.8	14.6	1317	34	US-09-910-186A-15

22 198.2 14.5 1317 17 US-09-350-756-5 Sequence 5, Appl1  
23 169.6 12.4 1351 1 PCT-US97-15394-35 Sequence 35, Appl  
24 169.6 12.4 1351 11 US-08-704-159-35 Sequence 35, Appl  
25 169.6 12.4 1351 43 US-10-271-012-35 Sequence 35, Appl  
26 169.4 12.4 1323 23 US-09-611-419A-3 Sequence 3, Appl1  
27 169.4 12.4 1323 34 US-09-910-186A-3 Sequence 3, Appl1  
28 169.4 12.4 1402 1 PCT-US97-15394-25 Sequence 25, Appl  
29 169.4 12.4 1402 8 US-08-405-496-25 Sequence 25, Appl  
30 169.4 12.4 1402 8 US-08-422-711-25 Sequence 25, Appl  
31 169.4 12.4 1402 8 US-08-480-604-25 Sequence 25, Appl  
32 169.4 12.4 1402 11 US-08-704-159-25 Sequence 25, Appl  
33 169.4 12.4 1402 14 US-09-084-517-25 Sequence 25, Appl  
34 169.4 12.4 1402 43 US-10-271-012-25 Sequence 25, Appl  
35 168.8 12.3 1326 23 US-09-611-419A-5 Sequence 5, Appl1  
36 168.8 12.3 1326 34 US-09-910-186A-5 Sequence 5, Appl1  
37 168.8 12.3 1330 1 PCT-US97-15394-22 Sequence 22, Appl  
38 168.8 12.3 1330 8 US-08-405-496-22 Sequence 22, Appl  
39 168.8 12.3 1330 8 US-08-422-711-22 Sequence 22, Appl  
40 168.8 12.3 1330 8 US-08-480-604-22 Sequence 22, Appl  
41 168.8 12.3 1330 11 US-08-704-159-22 Sequence 22, Appl  
42 168.8 12.3 1330 14 US-09-084-517-22 Sequence 22, Appl  
43 168.8 12.3 1330 43 US-10-271-012-22 Sequence 22, Appl  
44 168.8 12.3 1332 23 US-09-611-419A-1 Sequence 1, Appl1  
45 168.8 12.3 1332 34 US-09-910-186A-1 Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-611-419A-9  
; Sequence 9, Application US/09611419A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Leonard A.  
; APPLICANT: Byrne, Michael P.  
; APPLICANT: Middlebrook, John L.  
; APPLICANT: Lapenotiere, Hugh  
; APPLICANT: Clayton, Michael A.  
; APPLICANT: Brown, Douglas R.  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; FILE REFERENCE: A33626 067252.0105  
; CURRENT APPLICATION NUMBER: US/09/611.419A  
; CURRENT FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: PCT/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/146,192  
; PRIOR FILING DATE: 1999-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic construct based on BONTA Hc  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10)...(1359)  
US-09-611-419A-9

Query Match 100.0%; Score 1371; DB 23; Length 1371;  
Best Local Similarity 100.0%; Pred. No. 3.8e-279;

Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAATTCAGGATGACCATCCCATTCACATCTTCTCTACACCAACACTCCCTGTTGAAG 60  
Db 1 GAATTCAGGATGACCATCCCATTCACATCTTCTCTACACCAACACTCCCTGTTGAAG 60  
Qy 61 GACATCATCAACGAGTACTTCAACACATCAACGACTCCAGATCTCTGTCCTGTCGAGA 120  
Db 61 GACATCATCAACGAGTACTTCAACACATCAACGACTCCAGATCTCTGTCCTGTCGAGA 120  
Qy 121 CGTAGAACACCTTGGTCGACACCTCCGTTACACCCGAGGTCTCCGAGGAGGTGAC 180  
Db 121 CGTAGAACACCTTGGTCGACACCTCCGTTACACCCGAGGTCTCCGAGGAGGTGAC 180  
Qy 181 GTCCAGCTGAACCCCAATCTTCCCATTCGACTTCAAGTGGGTCTCCCGGTGAGGACAG 240  
Db 181 GTCCAGCTGAACCCCAATCTTCCCATTCGACTTCAAGTGGGTCTCCCGGTGAGGACAG 240  
Qy 241 GGTAGGTCATCGTCACCCAGACGAGAAATCGTCTACAACTCCATGTACGAGTCTCTTC 300  
Db 241 GGTAGGTCATCGTCACCCAGACGAGAAATCGTCTACAACTCCATGTACGAGTCTCTTC 300  
Qy 301 TCCATCTCTTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATC 360  
Db 301 TCCATCTCTTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATC 360  
Qy 361 ATCGACTCCGTCAGAACAACTCCGTTGGTTCATCGGTATCATCTCCAACTTCTCTGGTC 420  
Db 361 ATCGACTCCGTCAGAACAACTCCGTTGGTTCATCGGTATCATCTCCAACTTCTCTGGTC 420  
Qy 421 TTCACCTTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCC 480  
Db 421 TTCACCTTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCC 480  
Qy 481 AACACGCTCTCTGTTACAAAGTGGTCTTCTGTCACCGTCACCAACACATGATGGGT 540  
Db 481 AACACGCTCTCTGTTACAAAGTGGTCTTCTGTCACCGTCACCAACACATGATGGGT 540  
Qy 541 AACATGAAGATCTACATCAACGGTAAGCTGATGACACCATCAAGTCAAGGAGTGAAC 600  
Db 541 AACATGAAGATCTACATCAACGGTAAGCTGATGACACCATCAAGTCAAGGAGTGAAC 600  
Qy 601 GGTATCAACTCTCCAGACCATCACTTCGAGATCAACAGATCCAGACACCGGTCTG 660  
Db 601 GGTATCAACTCTCCAGACCATCACTTCGAGATCAACAGATCCAGACACCGGTCTG 660  
Qy 661 ATCACTCCGACTCCGACCAACATCAACATGTGGATCGGTACTTACATCTTCGCCAAG 720  
Db 661 ATCACTCCGACTCCGACCAACATCAACATGTGGATCGGTACTTACATCTTCGCCAAG 720  
Qy 721 GAGTTGACGATAGGACATCAACATCTCTCAACTCCTTGGAGTACACCAACGTCGTC 780  
Db 721 GAGTTGACGATAGGACATCAACATCTCTCAACTCCTTGGAGTACACCAACGTCGTC 780  
Qy 781 AAGGACTCTGGGTAAAGGATCAACATCTCTCAACTCCTTGGAGTACACCAACGTCGTC 840  
Db 781 AAGGACTCTGGGTAAAGGATCAACATCTCTCAACTCCTTGGAGTACACCAACGTCGTC 840  
Qy 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACACGAGTAAC 900  
Db 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACACGAGTAAC 900  
Qy 901 AACACGACTTCAACGAGGTACAGATCATCATCAGGTATCAGAGGTATACACCAAC 960  
Db 901 AACACGACTTCAACGAGGTACAGATCATCATCAGGTATCAGAGGTATACACCAAC 960  
Qy 961 GACACGAGTCCAGAGGTGGTACATCTCTTCTGACATGACTATCAACAAAGGCC 1020  
Db 961 GACACGAGTCCAGAGGTGGTACATCTCTTCTGACATGACTATCAACAAAGGCC 1020  
Qy 1021 TACAACCTGTTTCAAGAACGAGACCATGTACGCCGACACCACTCCACCGAGGATC 1080  
Db 1021 TACAACCTGTTTCAAGAACGAGACCATGTACGCCGACACCACTCCACCGAGGATC 1080









STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/704,159  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40, 027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 703-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3873  
US-08-704-159-59

Query Match 48.5%; Score 565.2; DB 11; Length 3876;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATCAACATCTCTCCACACCAACACTCCCTGTTGAAGGACATCATCAAC 72  
DB 2527 ACATACCCCTTAAATATTTTCATATCTACTATAATCTTTTAAAGATATAATTAAT 2586  
QY 73 GAGTACTTCAACACATCAACGACTCCAGATCCCGTCCCTGCGAGACCGTAAGAACCC 132  
DB 2587 GAATATTTCAATTAATTAATGATTCAAAATTTTGAGCCCTACAAACAGAAAAATACT 2646  
QY 133 TTGGTCGACACCTCCGTTTACAAACCCGAGGTCCTCGAGGAGGTGAGTCCACCTGAAC 192  
DB 2647 TTAGTGATACATCAGGATATTAATCGAAGAGTGTGAGAGCGGATGTCAGCTTAAT 2706  
QY 193 CCAATCTCCCATTCGACTTCAAGCTGGGTTCTCCGCTGAGGACAGAGTAAGGTGATC 252  
DB 2707 CCAATATTTCCATTTGACTTTAAATTAGTGTTCAGGGGAGGATAGAGGTAAGTTATA 2766  
QY 253 GTACCCAGACGAGACATCTCTACACCTCCAGTACGAGTCTTCCATCTCCTTC 312  
DB 2767 GTAACCCAGAAATTAATTTATATTAATCTGTATGTAAGTTTATAGCATTAAGTTT 2826  
QY 313 TGCATCAGATCAACAAAGTGGTCTCCACTTGCAGGTTACACATCATCGACTCCGTC 372  
DB 2827 TGCATTAGTAATAAATAAGTGGTAAGTAATTTACCTGGATATCTATAAATGATAGTGT 2886  
QY 373 AGAACAACCTCCGTTGGTCCATCGGTATCATCTCCACTCCCTGGTCTTCCACCTGAG 432  
DB 2887 AAAATAACTCAGGTGGAGTAGTAGTATTAATTAATTTTATAGTATTTACTTTAAA 2946  
QY 433 CAGAAGCAGGACTCCGAGCAGTCCATCACTTCTCCTACGACATCTCCAAACAGCTCCT 492  
DB 2947 CAAATGAAGATAGTAACAAAGTATAAATTTAGTTATGATATATCAAAATATGCTCCT 3006  
QY 493 GGTACAAAGTGGTCTCTGTCACCCGTCACCAACACATGATGGTAACATGAGATC 552  
DB 3007 GATACAAATAAATGGTTTTTGTAACTGTGTACTAACAATATGATGGGAAATATGAAGAT 3066  
QY 553 TACATCAACGGTAAGTGTATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTC 612

DB 3067 TATATAAATGGAATAATTAATAGATCTATATAAAGTTAAAGAACTAACTGGAATTAATTTT 3126  
QY 613 TCCAAGACCATCACCCTTCGAGATCAACAAGATCCGACACCGCTCTGATCACTCCCGAC 672  
DB 3127 AGCAAACTATAACATTGGAATAAATAAATTCAGATACCGGTTTGATTTACTTCAGAT 3186  
QY 673 TCCGACACATCAACATGTGTGATCCGTTGACTTCTACATCTTCCCAAGGAGTTGGAGGT 732  
DB 3187 TCTGATACATCAATATGTGATAGAGATTTTATATATTTTCTAAAGAAATAGATGTT 3246  
QY 733 AAGGACATCAACATCCCTGTTCAACTCTTCAGTACACCAACGCTCGTCAAGGACTACTGG 792  
DB 3247 AAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3306  
QY 793 GGTAAAGGCTGAGATACACAAGGAGTACTACATGTGTCAACATCGACTACTTTGAACAGA 852  
DB 3307 GGAATGATTTAAGATATAATAAAGATATTAATGTTTAATAGATATTTAAATAGA 3366  
QY 853 TACATGTAGCCCACTCCAGACAGATCGTCTTCAACACCAAGAGTACACACAGCTTC 912  
DB 3367 TATATGTATGCAACTCAGGACAAATTTTAAATACAGTGAATAATAATGACTTC 3426  
QY 913 AACGAGGTTTACAAGATCATCATCAAGGATATCAGAGTAAACCAACGACACAGAGTC 972  
DB 3427 AATGAGGATTAATTAATTAATAAAGATATCAGAGGAAATACAAATGATAGTAGTA 3486  
QY 973 AGAGTGTGATCCTCTGATCTTCGACATGACTATCAACAACAGGCTTACACCTGTTTC 1032  
DB 3487 CGAGGAGGAGATATTTTATATTTTATGATGACAAATTAATAACAAAGCATATAATTTGTTT 3546  
QY 1033 ATGAAGAAGGAGACCATGTACGCGACAAACACTCCACCGAGGAGCATCTACGCCATCGGT 1092  
DB 3547 ATGAAGATGAACATGTATGTCAGATTAATCATAGTACTGAAGATATATATGATAGT 3606  
QY 1093 CTGCGTGAGCAGACCAAGGATCAACGACAAACATCATCTCCAGATCCGAGCCAAATGAAC 1152  
DB 3607 TTAAGAAGAACAAACAAAGGATATAAATGATTAATTAATTAATTAATTAATTAATTAAT 3666  
QY 1153 ACACCTTACTACTCGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212  
DB 3667 AATCTTATTAATGACATCTCAATATTTAAATCAAAATTTAATGGAGAAATATTTCT 3726  
QY 1213 GGTATCTGTCCATCGTACCTACAGATTCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1272  
DB 3727 GGAATATGTTCAATAGGTACTATCGTTTGTAGACTTGGAGGTGATTTGGTATAGACAAAT 3786  
QY 1273 TACTTGGTCCAACTGTCAAGCAGGTAACGCTCTTGTGGAGTCCACTTCCACTCCACC 1332  
DB 3787 TATTGGTGGCTTACTGTGAAGCAAGAAATTTATGCTTCATTTATAGATCAACATCAACT 3846  
QY 1333 CACTGGGATTCGTCCTCCAGTCTCCGAGTAA 1362  
DB 3847 CATTTGGGTTTGTACCTGTAAGTGAATAA 3876

RESULT 6  
US-10-205-516-5  
; Sequence 5, Application US/10205516  
; GENERAL INFORMATION:  
; APPLICANT: Zhong, Jun  
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
; FILE REFERENCE: J2BtX1  
; CURRENT APPLICATION NUMBER: US/10/205,516  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: Clostridium botulinum  
; FEATURE:  
; NAME/KEY: CDS

LOCATION: (1), (3876)  
PUBLICATION INFORMATION:  
AUTHORS: Hauser, D. F.  
TITLE: Organization of the botulinum neurotoxin C1 gene and  
TITLE: its associated non-toxic protein genes in Clostridium  
TITLE: botulinum C 468  
JOURNAL: Mol. Gen. Genet.  
VOLUME: 243  
ISSUE: 6  
PAGES: 631-640  
DATABASE ACCESSION NUMBER: GeneBank / X2793  
DATABASE ENTRY DATE: 1993-05-03  
RELEVANT RESIDUES: 1 TO 3876  
US-10-205-516-5

Query Match 48.5%; Score 665.2; DB 42; Length 3876;  
Best Local Similarity 68.3%; Fred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATCAACATCTCTCTACACCAACATCCCTGTTGAAGGACATCATCAAC 72  
Db 2527 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTTAAAGATATAATTAAT 2586

QY 73 GAGTACTTCACACATCAACAGATCCAGATCCCTGTCCTCGCAACCGTAAGACACC 132  
Db 2587 GAATATTCAATAATATTAATGATTCAAAATTTTGAGCCTACAAAACAGAAAATAACT 2646

QY 133 TTGGTCGACACCTCCCGTTACACGCGGAGGTCTCCGAGGAGGTGACGTCACGCTGAAC 192  
Db 2647 TTAGTGATACATCAGGATATATGCAAGAGTGAGTGAAGAGGCGATGTTTCAGCTTAAT 2706

QY 193 CCAATCTCCATTCACATCAAGCTGGGTCTCCGGTGAGGACAGAGGTAAGGTCAIC 252  
Db 2707 CCAATATTCATTTGACCTTAAATAGTAGTTCAGGAGGAGTAGAGGTAAGTTATA 2766

QY 253 GTACCCAGAACAGAACATCGTCTACAACTCCATGTACGAGTCCCTCTCCATCTCCCTC 312  
Db 2767 GTACCCAGAACAAATATGTATATATCTATGATGAAGTTTAGCATTAGTTT 2826

QY 313 TGGATCAGAAATCAACAGTGGTCTCCACTTCCAGGTTCACCATCATCGCTCCGTC 372  
Db 2827 TGGATTAGAATAAATAAGTGGTAAAGTAAATTTACCTGGATATATTAATGATAGTGT 2886

QY 373 AAGAACAACTCCGGTGGTCCATCGGTATCATCTCCAACTTCTCTGCTCTCCACCTGAAG 432  
Db 2887 AATAAATCACTGGTGGATATAGTATATTTAGTATTTTACCTTTAA 2946

QY 433 CAGAACGAGCTCCGAGAGTCCATCACTCTCTCTAGCATCTCCAAACACGCTCCT 492  
Db 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATATCAATAATGCTCCT 3006

QY 493 GGTACAAACAGTGGTCTTCGTCACCGTCACCAACACATGATGGTAAACATGAAGATC 552  
Db 3007 GGATACAAATAATGGTTTTGTACGTATACAAATATGATGGGAATGAAGAT 3066

QY 553 TACATCAACGGTAAAGTGTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTC 612  
Db 3067 TATATAAATGAAATTAATAGATATATAAAGTTAAAGAACTAACTGGAATTAATTT 3126

QY 613 TCCAAGACCTACCTCCAGATCAACAGATCCAGACACCGGTCTGATCACTCCGAC 672  
Db 3127 AGCAAACTATAACATTTGAATAAATAAATTTCCAGATACCGGTTGATCTACTTCAGAT 3186

QY 673 TCGGACAACATCAACATGGTGGATCCGTGACTTCTACATCTTCGCAAGGAGTTGACGGT 732  
Db 3187 TCTGATAACATCAATATGTTGATGAAGAGATTTTATATATTTCGTAAGAATTAGATGGT 3246

QY 733 AAGGACATCAACATCTGTTCACTCTCTGAGTACACCAACGCTGCTCAAGGACTACTGG 792  
Db 3247 AAGATATTAATATATTTAATAGCTTGAATATACATAATGTTGTAAGAGATTATTGG 3306

QY 793 GGTAAACGACCTGAGATACAAACAGGAGTACTACATGGTCAACATCGACTACTGTAACAGA 852

Db 3307 GGAAATGATTTAAGATATAAATAAGATATATATATGGTTAATATAGATTATTTAAATAGA 3366

QY 853 TACATGTACGCCCAACTCCAGACAGATCGTCTTCACACCCAGACGTAACACACGACTTC 912

Db 3367 TATATGTATCGGAACCTCAGCAAAATTTGTTTAAATACACGTAGAAATAAATAGACTTC 3426

QY 913 AAGGAGGGTTCAAGATCATCATCAAGCGTATCAGAGGTAAACCAACGACACACGAGTC 972

Db 3427 AATGAAGGATATAAATATATAAAGAAATCAGAGGAATAACAATGATPACTAGAGTA 3486

QY 973 AGAGTGGTGACATCCCTGCTACTTCGACATGACTATCAACAACAGGCTTACACCTGTTTC 1032

Db 3487 CGAGGAGGAGATATTTTATATTTTATGATGACATTAATAACAAGACATATAATTTGTTT 3546

QY 1033 ATGAACACGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT 1092

Db 3547 ATGAAGAAATGAACATGATGATCGACATATCATATGATGATGATATATGCTATAGGT 3606

QY 1093 CTCGCTGAGCAGCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152

Db 3607 TTAAGAGAAACAAACAAAGGATATAAATGATATATATTTTCAATACAAACCAATGAAT 3666

QY 1153 AACACATCTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGGTGAGAACATCTCC 1212

Db 3667 AATCTATATTTACGCTCTCAATATTTAATCAATTTTATGAGAAATATTTCT 3726

QY 1213 GSPATCTGTCCATCGGTACCTACGATTCGCTGCTGGGTGGTGTGCTGTGATGATGATGATGAT 1272

Db 3727 GGAATATGTTCAATAGTACTTATGCTTTAGACTTGGAGGTGATTTGGTATAGACAAAT 3786

QY 1273 TACTTGTTCCTCACTGTCAGCAGGAGTAACTACGCTCTTCTGCTGGAGTCCACTTCCACC 1332

Db 3787 TATTTGGTCCCTACGTGAGCAGGAAATTTAGTCTTATTATTAGAAATCAATCAATCACT 3846

QY 1333 CACTGGGAGTTCCTCCAGTCTCCGAGTAA 1362

Db 3847 CATTTGGGTTTGTACTCTGTAAGTGAATAA 3876

## RESULT 7

US-10-271-012-59  
Sequence 59, Application US/10271012  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
Botulinum Neurotoxin  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/10/271,012  
FILING DATE: 15-Oct-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,159  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 59:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3876 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..3873  
 SEQUENCE DESCRIPTION: SEQ ID NO: 59:  
 US-10-271-012-59

Query Match 48.58; Score 665.2; DB 43; Length 3876;  
 Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCAATCAACATCTTCTCTACCAACAACCTCCCTGTTGAGGACATCATCAAC 72  
 DB 2527 ACAATACCCITTAATATTTTTCATATACTAATAATTCITTTATAAGAGATATAAT 2586

QY 73 GAGTACTTCACACATCAACGATCCGATCTGTCCTCGAGACCGGTAGAACACC 132  
 DB 2587 GAATATTTCAATAATTAATGATTCAAAAATTTTGAGCCCTACAAACAGAAAAATCT 2646

QY 133 TTGGTCGACACCTCGGTTACAACGCCGAGCTCTCCGAGGAGGTGAGCCAGCTGAAC 192  
 DB 2647 TTAGTGGATACATCAGGATATTAATGAGAGTGTGAGTGAAGAGCGATGTTCACTTAA 2706

QY 193 COAATCTTCCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGAGGTAAAGTCATC 252  
 DB 2707 COAATATTTCCATTTGACTTTAAATAGTAGTGTGAGGAGGAGTGAAGTAAAGTTATA 2766

QY 253 GTACCCAGAGAGAGACATCGTCTACAACTCCGATGACGAGTCTCTCCATCTCCTTC 312  
 DB 2767 GTACCCAGAGTGAATATTTGATTAATTTCTATGATGAAGTTTATAGATTTT 2826

QY 313 TGGATCAGAATCAACAGTGGGTCTCCAACTGCGAGTTACACCATCACTGACCTCCGTC 372  
 DB 2827 TGGATAGGATTAATAATAGGTAAGTAAATTTTACCTGGATATCTATAATTTAGTGT 2886

QY 373 RAGAACACTCCGTTGTCTCATCGGTATCATCTCCAACTTCTGGTCTCACCTGAAG 432  
 DB 2887 AAAATAACTCAGGTGGAGTATAGTATATTTAGTAAATTTTATAGTATTTACTTTAA 2946

QY 433 CAGAACGAGGACTCCGACGATCCATCACTTCCTTACGACATCTCCACACACGCTCCT 492  
 DB 2947 CAAATGAAGATAGTGAACAAAGTAAATTTTATGATATATCAAAATATGCTCCT 3006

QY 493 GGTACACACAGTGGTCTTCTGTCACGTCACCAACACATGATGGTACATGAAGATC 552  
 DB 3007 GGATCAATAAATGTTTGTGACGTGTTACTAACATATGATGGGAATATGAAGATT 3066

QY 553 TACATCAACGTTAAGTCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCACTTC 612  
 DB 3067 TATATAATGGAATTAATAGATCTATAAAAGTTAAAGAACTAACTGGAATTAATTT 3126

QY 613 TCCAGACATCACTCCGATCGAGATCAACAGATCCGACACCGGTCTGATCACTCCGAC 672  
 DB 3127 AGCAAACTATAAATTTGAAATTAATAAATTTCCAGATACCGGTTGTTACTTCAGAT 3186

QY 673 TCCGACACATCAACATGCGATCGGTGACTTCTACATCTTCGCAAGGAGTTGGAGCGGT 732  
 DB 3187 TCTGATACATCAATATGCGATAAGATTTTATATATTTGCTAAGATATAGATGGT 3246

QY 733 RAGGACATCAACATCTCTTCACTCTTCACTGAGTACACCAACGTCGTCAGGACTACTGG 792  
 DB 3247 AAAGATATAATATTTTAAATAGTGTGAATATATCTAATGTTGTTAAAGATTTATGG 3306

QY 793 GGTACGACCTGAGATACAAAGGAGTACATGCTGACATGCTGACATCTACTTGAACAGA 852  
 DB 3307 GGAATGATTAAAGATATTAAGATATTTATATGTTTATATAGATTTTAAATAGA 3366

QY 853 TACATGACGCCAACTCCAGACAGATCGTCTTCAACACACGAGCTGTAACAAACGACTTC 912  
 DB 3367 TATATGTATCGAACTCAGCAAAATTTGTTTTTAATACACGCTAGAAATTAATATGACTTC 3426

QY 913 AACGAGGTTACAGATCATCAAGCGTATCAGAGGTAAACACCAACGACACACAGAGTC 972  
 DB 3427 AATGAAGATATAAATTAATAAAGAGATCAGAGGAATACAAATGATCTAGAGTA 3486

QY 973 AGAGTGTGACATCTCTGACTTCGACATGACTATCAACAAAGGCTTACACCTGTTTC 1032  
 DB 3487 CGAGGAGGATATTTTATATTTTATGATGACAAATTAATAACAAAGCATATAATTTGTT 3546

QY 1033 ATGAAGACGAGACCATGTAGCGCGCAACCACTCCACCGAGGACATCTACGCCATCGGT 1092  
 DB 3547 ATGAAGATGAATATGTATGAGATATCATATAGTACTGAGATATATATGCTATAGT 3606

QY 1093 CTGGTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCAATGAAC 1152  
 DB 3607 TTAAGAGAACAAACAAAGGATATAAATGATATATATTTTCAATACAAACCAATGAAT 3666

QY 1153 AACACTTACTACTACGTTCCGAGATCTTCAAGTCCAACTTCAACGCTGAGACATCTCC 1212  
 DB 3667 AATCTTATTTATGAGCATCTCAAAATTTTAAATCAAAATTTTATGGAGAAATATTTCT 3726

QY 1213 GGTATCTGTTCCATCGGTACTTACAGATTCGCTCTGGTGGTGGTACTGTTACAGACAA 1272  
 DB 3727 GGAATATGTTCAATAGTACTTATTCGTTTATAGACTTGGAGGTGATGGTATAGACAAAT 3786

QY 1273 TACTTGGTCCAACTGTCAGACGAGGTAACTAGCCCTCTTCTGCTGGAGTCCACTTCCACC 1332  
 DB 3787 TATTGTGCTACTGTGAACGAGGAATTTATGCTTCATTTATTAATCAACATCAACT 3846

QY 1333 CACTGGGATTCGTCGCCAGTCTCCGAGTAA 1362  
 DB 3847 CATGGGGTTTGTACCTGTAAAGTGAATAA 3876

RESULT 8  
 US-08-954-302-2  
 ; Sequence 2, Application US/08954302  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lance Simpson, Nikita Kiyatkin,  
 ; APPLICANT: Andrew Maksymowich  
 ; TITLE OF INVENTION: Compositions and Methods for Systemic  
 ; TITLE OF INVENTION: Delivery of Oral Vaccines and Therapeutic Agents  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Jane Massey Licata, Esq.  
 ; STREET: 66 E. Main Street  
 ; CITY: Marlton  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 08053  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
 ; COMPUTER: IBM 486  
 ; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/954,302  
 ; FILING DATE: herewith  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jane Massey Licata  
 ; REGISTRATION NUMBER: 32,257  
 ; REFERENCE/DOCKET NUMBER: JEFF-0164  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (609) 779-2400  
 ; TELEFAX: (609) 810-1454

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3950  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
ANTI-SENSE: NO  
US-08-954-302-2

Query Match 48.5%; Score 665.2; DB 13; Length 3950;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTTCCTCCACACCAACTCCCTGTTGAAGGACATCATCAAC 72  
Db ACAAATACCTTTAAATATTTTTCATATCTAATAATCTTTTATTAAGAGATATAATTAAT 2617  
QY 73 GACTACTTCAACACATCAACGACATCCAGATCCCTGTCGCAACCGTAAGAACACC 132  
Db GAATATTTCAATATATTAATGATCAAAAATTTGAGCCTACAAAACAGAAAATTAAT 2677  
QY 133 TTGGTCGACACCTCCCGTTACAAACCGGAGGTCTCCGAGGAGGTGACGTCCAGCTGAAC 192  
Db TTAGTGGATACATCAGGATATAATGCAAGGTGAGTGAAGAGGCGATGTTTCAGCTTAAT 2737  
QY 193 CCAATCTTCCATTCACATCAAGCTGGGTTCTCCGGTGAGGACAGAGGTAAAGTCAATC 252  
Db CCAATATTTCCATTTGACATTAATAGTAGTCTCAGGAGGAGTAGAGGTAAAGTTATA 2797  
QY 253 GTCAACCCAGAACGAGAACATCGCTTACAACTCCATGACGAGTCCCTTCCATCTCCCTTC 312  
Db GTAAACCAGAAATGTAATATTAATCTATGATGAAGATTTTAGCATTAATGTTT 2857  
QY 313 TGGATCAGATCAACAAAGTGGGTCCCACTTCGACAGGTACACCATCATCGACTCCGTC 372  
Db TGGATGAGAAATAAATAATGAGTAAATATTTACCTGGATATACATAATGATAGTGT 2917  
QY 373 AAGAACAACCTCCGTTGGTCCATCGGTATCATCTCCAACTTCCTGTGCTTCACCCGTGAAG 432  
Db AAAAAATACCTCAGGTGGAGTATAGTATATTAATAGTAATTTTATAGTATTTTAAAT 2977  
QY 433 CAGAACGAGACCTCCGAGACATCCACTTCTCCTAGGACATCTCCCAACACACCTCCT 492  
Db CAAATGAGATAGTGAACAAAGATATAATTTAGTATGATATATCAATAATGCTCCT 3037  
QY 493 GGTATCAACAAAGTGGTTCCTGTCACCGTCACCAACACATGATGGTAAACATGAAGATC 552  
Db GGATACAATAATGTTTGTGTAAGTGTACTTAAATATGATGGAAATATGAAGAT 3097  
QY 553 TACATCAACGGTAACTGATGACACCATCAGGTCAAGGATGACCGGTATCAACTTC 612  
Db TATATTAATGGAATTAATATAGATCTATAAAGTTAAAGAACTAACTGGAATTAATTT 3157  
QY 613 TCAAGACCATCACTTCGAGATCAACAGATCCCAAGACACCGGTCTGATCACTCCGAC 672  
Db AGCAAACTATACATTTGAATTAATAAATTCAGATACCGGTTGATTTACTTACAT 3217  
QY 673 TCCGACAAACATCAACATGAGTGGATCCGTGACTTCTACATCTTCGCCAAGGATTTGGACGGT 732  
Db TCTGATACATCAATATGTGGATAAGAGATTTTATATATTTTCTTAAAGAAATTAGATGGT 3277  
QY 733 AAGGACATCAACTCCTGTTCAACTCTTCGAGTACACCAACGCTGCTCAAGGACTACTGG 792  
Db AAGATATTAATATATTAATTAATAGCTTGCAATATACATATGTTGTAAAGATTAATGG 3337  
QY 793 GGTAACGACCTGAGATACAAACAGGAGTACTACATGGTCAACATCGACTTACTTGAACAGA 852  
Db GGAAATGATTTAAGATATAATAAAGAAATATTAATGTTAATATAGATTTATTAATAGA 3397  
QY 853 TAGATGTACGCCAACCTCCAGACAGATCGCTTCAACACACGAGGTAAACACACGACTTC 912  
Db TATATGTATCGGAACCTCACAGAAATGTTTAAATACAGTAGGAAATATAATGACTTC 3457

QY 913 AACGAGGTTACAGATCATCATCAAGCTATCAGAGGTACACCAACGACACACGAGATC 972  
Db AATGAGGATATAAAATTAATAAAGAAATCAGAGGAAATACAAATGATAGAGTA 3517  
QY 973 AGAGGTGGTGACATCCTGTACTTCGACATGACTATCAACACAGGCTTACACCTCTTC 1032  
Db CGAGGAGGAGATATTTTATATTTTATGATGACAATTAATAAAGAGATATAATTTGTT 3577  
QY 1033 ATGAAGAACGAGACCATGTACGCCGACCAACCACTCCACGAGGACATCTAGCCATCGGT 1092  
Db ATGAAGATGAACATGTATGTCAGATATATCATAGTACTAGAGATATATATCTATAGGT 3637  
QY 1093 CTGCGTGACCAACCAAGGACATCAACGACACATCATCTTCCAGATCCGACCAATGAC 1152  
Db TTAAGAGAACCAACAAAGGATATAAATGATATAATATATATTTCAATACAAACCAATGAAT 3697  
QY 1153 ACACATTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGTTGAGAACATCTCC 1212  
Db AATACTTATTAATACGATCTCAATATTTAAATCAAAATTTAATGGAGAAATATTTCT 3757  
QY 1213 GGTATCTGTTCATCGGTACCTACAGATTCGCTGGGTGGTGGTACTGGTACAGACAAAC 1272  
Db GGAATATGTTCAATAGGTACTTATCGTTTACACTTGGAGGTGATTTGATAGACACAAT 3817  
QY 1273 TACTTGGTTCCTCAACTGTCAAGCAGGTAACTACGCTCCTTCTGGAGTCCACTTCCACC 1332  
Db TATTTGGTGCCTACTGTGAAGCAAGAAATTAATGCTTCATTTAATGAATCAACATCACT 3877  
QY 1333 CACTGGGATTCGTCCTCCAGTCTCCGAGTAA 1362  
Db CATTTGGGTTTGTACCTGTAAGTGAATAA 3907

## RESULT 9

US-10-205-516-19  
Sequence 19, Application US/10205516  
GENERAL INFORMATION:

APPLICANT: Zhong, Jun  
TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
TOXIN: Neurotoxins Through Recombinant DNA Technique  
FILE REFERENCE: JZbxc1  
CURRENT APPLICATION NUMBER: US/10/205,516  
CURRENT FILING DATE: 2002-07-25  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 19  
LENGTH: 3906  
TYPE: DNA  
ORGANISM: Clostridium botulinum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3906)  
FEATURE:  
NAME/KEY: misc\_signal  
LOCATION: (1347)..(1358)  
OTHER INFORMATION: factor Xa site  
FEATURE:  
NAME/KEY: misc\_signal  
LOCATION: (3886)..(3903)  
OTHER INFORMATION: 6-histidine tag  
US-10-205-516-19

Query Match 48.4%; Score 663.2; DB 42; Length 3906;  
Best Local Similarity 68.4%; Pred. No. 1.2e-129;  
Matches 917; Conservative 0; Mismatches 423; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTTCCTTACACCAACCAACTCCCTGTTGAAGGACATCATCAAC 72  
Db ACAAATACCTTTAAATATTTTTCATATCTAATAATCTTTTATTAAGAGATATAATTAAT 2598  
QY 73 GAGTACTTCAACACATCAACGACATCCAGATCCCTGCTCCGACGAGTAAAGAACACC 132  
Db GAATATTTCAATATAATTAATGATTCAAAAATTTTGAGCCTACAAAACAGAAAAATCT 2658



QY 133 TTGTCGACACCTCCGGTTTCAACGCGGAGGTCTCCGAGGAGGTGACGTCGAGCTGAAC 192  
DB 2659 TTAGTGGATACATCAGGATATAATGCAAGAGTGAAGAGGCGATGTCAGCTTAAT 2718  
QY 193 CCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTGAGGACAGAGTGAAGTCATC 252  
DB 2719 CCAATATTCCTTGAATTAATAGGTAGTTCAGGGGAGGATAGAGTTAAGTTATA 2778  
QY 253 GTCACCCAGACGAGAACATCGCTTCAACTCCATCGATGAGGTCCTTCCATCCCTTC 312  
DB 2779 GTACCCAGAGTGAATAATTTGATATATAATTTCTATGATGAAAGTTTACCATTTAGTTT 2838  
QY 313 TGGATCAGATCAACAAGTGGGTCCTCAACTGCGAGGTACACCATCATCGACTCCGTC 372  
DB 2839 TGGATTAAGTAATAAATGGGTAAGTAATTTACCTGGATATCTATTAATTTAGTGT 2898  
QY 373 AAGAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTGCTTCCACCTGAAG 432  
DB 2899 AAAATAACTCAGGTTGGAGTATAGGTATTTAGTAATTTTGTAGTATTTTACTTTAAA 2958  
QY 433 CAGACGAGGACTCCGAGCAGTCCATCAACTTCTCCAGACATCTCCACACAGCTCCT 492  
DB 2959 CAAATGAAGTATGTAACAAAGTAAATTTAGTTATGATATCAATATATGCTCCT 3018  
QY 493 GTGTACAACAAGTGGTCTTCTGTCACCGTCACCAACAACATGATGGTTAACATGAAGTC 552  
DB 3019 GGATACAATAAATGGTCTTGTAACTGTTTACTTAACAATATGATGGAAATATGAAGAT 3078  
QY 553 TACATCAACGCTAGCTGACGACACCATCAAGTCAAGGTGACCGGTATCAACTTC 612  
DB 3079 TATATAAGTGAATAATTAAGATATCTAAAGTTAAAGAACTACTGGATTAATTT 3138  
QY 613 TCCAGACCATCACTCCGAGATCAACAAGATCCGACACACCGCTGTGATCACTCCGAC 672  
DB 3139 AGCAAACTATAACATTTGAATAAATAAATAATCCAGATACCGGTTGTACTTACGAT 3198  
QY 673 TCCGACACATCAACATGTGGATCCGTCGACTTCTACATCTTCCCAAGGATGGACGT 732  
DB 3199 TCTGATAACATCAATATGTGGATGAAGATTTTATATATTTCTTAAGAAATAGATGT 3258  
QY 733 AAGGACATCAACTCCCTGTTCACTCTCCAGTACACCAACGCTGTCAGGACTACTGG 792  
DB 3259 AAAGATATTAATATATTTAATAGCTTCAATATATCTGTAAGAGTATTTGG 3318  
QY 793 GGTACGACCTGACATACACAGGAGTACTACATGGTCAATCTGACATCTGACATGA 852  
DB 3319 GGAATGATTTAAGATATAATAAGAAATATATATGTTTAATATGATTTATTAATAGA 3378  
QY 853 TACATGTACGCCAACTCCAGACAGATCGTCTTCAACACACGAGGTAAACAACAGACTTC 912  
DB 3379 TATATGTATGGAACCTCAGACAAATTTGTTTTAATACAGGTAGATAATTAATGACTTC 3438  
QY 913 AACGAGGTTTACAGATCATCTCAAGCGTATCAGAGGTATACACCAACGACACAGACTC 972  
DB 3439 AATGAAGATATAAATAATATAAAGAAATCAGAGGAATACAAATGATATCTAGAGTA 3498  
QY 973 AGAGTGTGACATCCCTGTTACTTTCGACATGACTATCAACAACAGGCTTACACCTGTTC 1032  
DB 3499 CAGGAGGAGATATTTATATTTTGTATGATGACAATTAATACAAAGCATATATTTGTT 3558  
QY 1033 ATGAAGACGAGACCATGTACGGGAGCAACCATCTCCACGAGGACATCTACGCCATCGGT 1092  
DB 3559 ATGAAGATGAACATATGTATGCGAGATAATCATAGTACTGAAGATATATATGCTATAGT 3618  
QY 1093 CTGGTGAGCAGACCAAGGACATCAAGGACACATCATCTTCCAGATCCAGCAATGAAC 1152  
DB 3619 TTAGAGACACAAACAAGGATATAAGTAAATTAATTTTCAATACCAACCAATGAT 3678  
QY 1153 AACACTTACTACTACGTTTCCAGATCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212  
DB 3679 AATACTTATTATACGATCTCAATATTTTAAATCAAAATTTTAAATGGAGAAAATATTTCT 3738

QY 1313 GGTATCTGTTCCATCGGTACCTACAGTTCCGTCCTGGTGGTGTGACTGGTACAGACACAC 1272  
DB 3739 GGAATATGTTCAATAGGTACTTATCGTTTGTAGACTTTGGAGGTGATTTGGTATAGACACAAT 3798  
QY 1273 TACTTGGTTCCAACTGTCAGCAGGTAACACTAGCCTCCCTGCTGAGTCCACTTCCACC 1332  
DB 3799 TATTTGGTGGCTACTGTGAGCAAGGAATATGCTTCATTATTAGAAATCAACATCAACT 3858  
QY 1333 CACTGGGGATTCGTCOCAGT 1352  
DB 3859 CATTGGGGTTTGTACCTGT 3878  
RESULT 10  
US-09-730-237-4  
; Sequence 4, Application US/09730237  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Treating Hyperhidrosis  
; FILE REFERENCE: 2933  
; CURRENT APPLICATION NUMBER: US/09/730,237  
; CURRENT FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: botulinum toxin  
US-09-730-237-4  
Query Match 48.38; Score 662; DB 29; Length 3876;  
Best Local Similarity 68.1%; Pred. No. 2.2e-129;  
Matches 920; Conservative 0; Mismatches 430; Indels 0; Gaps 0;  
QY 13 ACCATCCCATTCACATCTCTCTCTACACCAACAACCTCCTGTTGAAGGACATCATCAAC 72  
DB 2527 ACATACCCCTTAATATTTTTCATATCTACTAATAATCTTTATTAAGATATATAAT 2586  
QY 73 GAGTACTTCAACAACATCAACGACTCCAGAGTCTCCCTGCTCCGACAGACCGTAAGAACACC 132  
DB 2587 GAATATTTCAATATAATTAATGATCAAAAATTTTGAAGCTTACAAAACAGAAAAATACT 2646  
QY 133 TTGTCACACACTCCCGGTTACAACGCGGAGGTCTCCGAGGAGGTGACGTCGAGCTGAAC 192  
DB 2647 TTAGTGATACATCAGGATATAATGACAGAGTGAAGAGGCGATGTCAGCTTAAT 2706  
QY 193 CCAATCTTCCATTCGACTTCAAGCTCGGTTCCTCCGGTGAGGACAGAGTGAAGTCATC 252  
DB 2707 CCAATATTTCCATTTGACTTTAAATAGTGTAGTTCAGGGGAGGATAGAGTAAAGTTATA 2765  
QY 253 GTCACCCAGACGAGACATCGTCTCAACTCCATCTAGAGTCTCTTCCATCTCCTTC 312  
DB 2767 GTACCCAGATGAATAATTTGTATATATTTCTATGATGAAGATTTTAGCATTTAGTTT 2826  
QY 313 TGGATCAGATCAACAAGTGGGTCCTCAACTTCCAGGTTACACCATCATCGACTCCGCTC 372  
DB 2827 TGGATTAAGATAAATAAATGGTAAAGTAAATTTTACCTGGATATCTACTAATAATGATAGTGT 2885  
QY 373 AAGAACAACTCCGTTGGTCCATCGGTATCATCTCCAACTCCTGCTTCCACCTGAAG 432  
DB 2887 AAAATAACTCAGTGGAGTATAGGTATTTAGTAATTTTGTAGTATTTTACTTTAAA 2946  
QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCTCCCAACACGCTCCT 492  
DB 2947 CAAATGAAGATAGTGAACAAGTATAAATTTTAGTTATGATATATCAATAATGCTCCT 3006  
QY 493 GGTATCACAAAGTGGTCTTCTGTCACCGTCACCAACAACATCATGGTAAACATGAAGTC 552  
DB 3007 GGATAGAAATAAATGGTGTGTTTGTAACTGTTACTAACAATATGATGGAAATATGAAGAT 3066  
QY 553 TACATCAACGGTAAAGTGAATCGACACCATCAAGGTCAAGAGTTGACCGGTATCAACTTC 612  
DB 3067 TATATAATGGAAAAATTAATAGATATCTATAAAGTTTAAAGAACTAACTGGAATTAATTTT 3126



QY	613	TCCAAGACCATCACCTTCGAGATCAACAAAGATCCAGACACCGGTCTGTGATCACTCCGAC	672
Db	3127	AGCAAACTATAACATTTGAAATAAATAAATTCAGATACCGGTTTGATTACTTCAGAT	3186
QY	673	TCCGACCAACATCAACATGNGATCGGTACACTCTTCACATCTTCGCCAAGGAGTTGGAGGT	732
Db	3187	TCGTATACATCATATATGCGATTAAGAGATTTTATATATTTGCTAAGAATTAGATGGT	3246
QY	733	AAGGACATCAACATCCTGTTCAACTCCTTGCAGTACACCACGTCGTCAGGACTACTG	792
Db	3247	AAAGATTAATAATATATTATTAATAGCTTCGAATATACTAATGTGTGTAAAGATTATTTGG	3306
QY	793	GSFAACGACCTGAGATACAAACAAAGGAGTACTACATGCTACACATCGACTACTTTGAACAGA	852
Db	3307	GGAAATGATTTAAGATATAATAAAGAAATATTATATGTTTATAATAGATTATTTAAATAGA	3366
QY	853	TACATGTACGCCAACTCCAGACAGATCGTCTTCAACACGACGACTACACAACAGCACTTC	912
Db	3367	TATATGTATGCGAACTCACGACAAATTTGTTTTTAACACGTAAGAAATAATAATGACTTC	3426
QY	913	AACGAGGTTACAGATCAATCATCAAGCGTATCAGAGGTACACCAACGACACACGAGTC	972
Db	3427	AATGAAGGATATAAAATTTAATAAAGAAATCAGAGAAATACAAATGATACTAGATA	3486
QY	973	AGAGGTGGTGACATCGCTGTACTTCGACATGACTATCAACAACAGGCGCTACACACCTGTT	1032
Db	3487	CGAGGAGGAGATATTTTATATTTTGNATGACAATTATAACAAGCATATATTTTGT	3546
QY	1033	ATGAAGAAGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT	1092
Db	3547	ATGAAGAAATGAAACTATGTATCGAATAATCATAGTACTGAAGATATATATGCTATAGGT	3606
QY	1093	CTGCGTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCAAATGAAC	1152
Db	3607	TTAAGAGAACAAACAAGGATATAATGATATATATATTTTCAATACAAACCAATGAAT	3666
QY	1153	AACACTTACTACTACGCTTCCGAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC	1212
Db	3667	AATACTTATTTATGSCATCTCAAATATTTAAATCAAATTTTAAATGAGAAATATTTCT	3726
QY	1213	GGTATCTGTTCCATCGGTACCTACAGATTCGGTCTGGGTGTGTGACTGGTACGACACAAC	1272
Db	3727	GGAAATGTTTCAATAGGTACTTATCGTTTTAGACTTGGAGGTGATTGGTATAGACACAAT	3786
QY	1273	TACTTGGTTCCAACTGTCAAGCAGGTAACATCGGCTCCTTTCGTGGAGTCCACTTCCACC	1332
Db	3787	TATTTGGTGCCTACTGTGAGCAAGAAATATATGCTTCATTTTAGAATCAACATCACT	3846
QY	1333	CATCGGGATTCGTCCCACTCTCCGAGTAA	1362
Db	3847	CATTGGGGTTTTGTACCTGTAAGTGAATAA	3876

RESULT 11	
US-10-051-952-4	
:	: Sequence 4, Application US/10051952
:	: GENERAL INFORMATION:
:	: APPLICANT: Walker, Patricia
:	: TITLE OF INVENTION: Methods of Administering Botulinum Toxin
:	: FILE REFERENCE: 2933CIP
:	: CURRENT APPLICATION NUMBER: US/10/051,952
:	: CURRENT FILING DATE: 2002-01-17
:	: PRIOR APPLICATION NUMBER: 09/730,237
:	: PRIOR FILING DATE: 2000-12-05
:	: NUMBER OF SEQ ID NOS: 12
:	: SOFTWARE: PatentIn Ver. 2.1
:	: SEQ ID NO 4
:	: LENGTH: 3876
:	: TYPE: DNA
:	: ORGANISM: botulinum toxin
US-10-051-952-4	
Db	3247 AAAGATATTATAATTATTAAATAGCTTGCATATATTAATGTTGTGTTAAAGAATTATTGG
Qy	793 GGTAAACGACCTGAGATCAACAAGGAGTACTACATGTCACATCGACTACTTGACACAGA
Db	3307 GGAATGATTTAAGATATAATAAGAATATTATATGTTAATATAGATTATTATTAAATAGA
Qy	853 TACATGTACGCCAACTCCAGACGATCGTCTTCAACACCACAGACGTAAACAACACGACTTC
Db	3367 TATATGTATGCGAACTCACGACAAAATTTGTTTTTAATACACGTAGAAATAATAATGACTTC
Qy	913 AACGAGGGTTTACAAGATCATCATCAAGCGTATCAGAGGTAAACCAACACGACACAGAGTC
Db	3427 AATGAAGGATATAAAATTTAATAAAAAAGGATCAGAGGAAATCAAATGATGATACTAGAGTA
Qy	973 AGAGTGGTGACATCCCTGCTACTTCGACATGACTATCAACAACAGGGCTACACACCTGTTTC
Db	3487 CGAGGAGGAGATATTTTATATTTTGTATGACAAATTAATAACAAGGACATATAATTGTTT
Qy	1033 ATGAGAACGAGACCATGTACGCGCAACCAACCTCCACGAGGACATCTACGCCATCGGT



Db 1255 ATAATATTATATTCAATAACACCAATGAATAATACTTATTATTACGCACTCAAAAT 1314  
 QY 1181 TCAAGTCCAACTTCAACGGTGAGAACATCTCCGGTATCTGTTCCATCGGTACTACAGAT 1240  
 Db 1315 TTAATCAATTTTATGAGAGAAATATTTCGGAATATGTTCAATAGGTACTTATCGTT 1374  
 QY 1241 TCGGTCTGGGTGACTGTGTACAGACACAACTACTTGTTCCTCAACTGTCAAGAGGTA 1300  
 Db 1375 TTAGACTGTGAGGTGATGTTGATAGACAAATATTATTGTTGCTACTGTGAAGCAAGGAA 1434  
 QY 1301 ACTACGGCTCTTGTGGAGTCCACTTCCACCACCTGCGGATTCGCCAGTCTCCGAGT 1360  
 Db 1435 ATATGCTTCATTATTAGATCAACATCACTCTTGGGTTTGTACCTGTAAGTGAT 1494  
 QY 1361 AATAG 1365  
 Db 1495 AAAAG 1499

## RESULT 13

US-08-704-159-61  
 ; Sequence 61, Application US/08704159  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williams, James A.  
 ; APPLICANT: Thalley, Bruce S.  
 ; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
 ; TITLE OF INVENTION: Botulinum Neurotoxin  
 ; NUMBER OF SEQUENCES: 82  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medlen & Carroll  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/704,159  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ingolia, Diane E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPD-02304  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 61:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1502 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 108..1493  
 US-08-704-159-61

Query Match 47.4%; Score 649.8; DB 11; Length 1502;  
 Best Local Similarity 68.2%; Pred. No. 7.1e-127;  
 Matches 903; Conservative 0; Mismatches 422; Indels 0; Gaps 0;

QY 41 CCAACAACTCCCTGTTGAAGGACATCATCAACGAGTCTTCAACACATCAACGACTCCA 100  
 Db 175 CTAGCATGGCTTTATTAAGAGATATATTAAGATATTTCAATATATATTAATGATCAA 234  
 QY 101 AGATCTGTCTCCGCAACCGGTGAAGAACCTTGTGTGACACCTCCGGTTACACGCG 160  
 Db 1315 TTAATCAATTTTAAATGGAGAAATATTTCGGAATATTTTCGGAATATGTTCAATAGTACTTATCGTT 1374

Db 235 AAAATTTTGGCCCTACAAAACAGAAAAAATACTTTAGTGGATACATCAGGATATATATGCG 294  
 QY 161 AGTCTCCGAGGAGGTGAGTCCAGCTCAACCCCAATCTTCCATTCGACTTCAAGCTGG 220  
 Db 295 AAGTGAGTGAAGAGCGGTGTTGAGCTTAATCCCAATATTCCATTTGACTTTAAATAG 354  
 QY 221 GTTCTCCGGTGAAGGACAGAGGTAAAGTTCATCGTCAACCCAGAACGAGAACATCGTCTCA 280  
 Db 355 GTAGTTTCAGGGGAGGATAGAGGTAAAGTTATAGTAACCCAGAAATGAAATATGTTATATA 414  
 QY 281 ACTCCATGTACGAGTCCCTCTCCATCTCTCTGGATCAGATCAACAGTGGGCTCCA 340  
 Db 415 ATTCTATGTATGAAGGTTTAGCATTTAGTTTGGATAGAATAAATAATGGTAAAGTA 474  
 QY 341 ACTTGCCAGGTTTACACCATCATCGACTCCGTCGAAGAACAACCTCCGTTGGTCCATCGGTA 400  
 Db 475 ATTTACCTGGATATATAATTTAGTTTAAATAAATCACTCAGTTGGAGTATAGTA 534  
 QY 401 TCACTCCAACTTCCCTGGTCTTCAACCTGAAGCAGACGAGGACTCCGAGGAGTCCATCA 460  
 Db 535 TTATTAGTAAATTTTGTAGTATTACTTTTAAACAAAATGAAGATAGTAGAACAAAGTATAA 594  
 QY 461 ACTTCTCCTACGACATCTCCAACACGCTCCCTGGTTTCAACAAGTGGTCTTCTCGTCAACG 520  
 Db 595 ATTTAGTTATGATATATCAATAATGCTCTCTGGATCAATAAATGGTTTTTGTAACTG 654  
 QY 521 TCACCAACACATGATGGGTAAATCATGATCATCAACGCTAAGCTGATCGACACCA 580  
 Db 655 TTACTAACATATGATGGGAAATATGAGATTTATATAATGAAATTAATAGTACTA 714  
 QY 581 TCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCAAGACCATCACTTCGAGATCAACA 640  
 Db 715 TAAAGTTAAAGAACTAAGTGAATTAATTTAGCAAACTATACATTTGAAATAAATA 774  
 QY 641 AGATCCAGACACCGGTCTGTGATCACTCCGATCCGACAAATCAACATGTGGATCCGTG 700  
 Db 775 AAAATCCAGATACCGGTTTGATTTACTTTCAGATTTCGATAACATCAATATGTGGATAAG 834  
 QY 701 ACTTCTACATCTTCGCCAAGGAGTTGGCGGTAAGGACATCAACATCTGTTCAACTCT 760  
 Db 835 ATTTTATATATTTGCTAAAGAAATTAGATGTTAAGATATTAATATATTTAATAGCT 894  
 QY 761 TGCAGTACACCAACGCTCGTCAAGGACTACTGGGGTAAAGACCTGAGATACAAAGAGGT 820  
 Db 895 TGAATATACTAATTTGTAAGAAATTTATGGGAAATGATTAAAGATATAATAAAGAT 954  
 QY 821 ACTACATGGTCAACATCGACTACTTGAACAGATACATGTACCCCACTCCAGACAGATCG 880  
 Db 955 ATTATATGGTTAATATAGATTATTTAAATAGATATATGTATCGAACTCAGACAAATG 1014  
 QY 881 TCTTCAACACCGACGTAAACAACAGACTTCAACGAGGGTTTCAAGATCATCATCAAGC 940  
 Db 1015 TTTTAAATACACGTAGAATAATAATGACTTCAATGAAGGATATAAATAATAAATAA 1074  
 QY 941 GTATCAGAGGTAAACCAACGACACGAGTCAAGGTGGTGCATCTGTTACTTCGACA 1000  
 Db 1075 GAATCAGAGGAAATACAAATGATAGAGTACGAGGAGGAGATATTTTATATTTTGATA 1134  
 QY 1001 TGACTATCAACAAACAGGCTCAACCTGTTTCATGAAGAACGAGACCATGTACGCGGACA 1060  
 Db 1135 TGACAATTAATACAAAGCATATAATTTCTTTATGAAGAAATCAAACTATGTATGAGATA 1194  
 QY 1061 ACCATCCACCGAGGACATCTAGCCCATCGGTCTCGGTGAGCAGACCAAGGACATCAAGC 1120  
 Db 1195 ATCATAGTACTGAAGATATATATGCTATAGTTTAAAGAAACAAACAAAGGATATAAATG 1254  
 QY 1121 ACAACATCATCTTCCAGATCCAGCCCAATGAACAACACTTACTACTACGCTTCCAGATCT 1180  
 Db 1255 ATAAATATATTTCAAAATACACACCATGAATATATTTATTTACGCACTCAATAT 1314  
 QY 1181 TCAAGTCCAACTTCAACGGGTGAAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGAT 1240  
 Db 1315 TTAATCAATTTTAAATGGAGAAATATTTCGGAATATTTTCGGAATATGTTCAATAGTACTTATCGTT 1374



QY 1241 TCGTGTGGGTGAGTGTGACAGACAACTACTTGTTCCTCAACTGTCAAGCAGGTA 1300  
Db 1375 TTAGACTTGGAGTGATGATGATAGACAAATATTGTGCTACTGTGAAGCAAGAA 1434  
QY 1301 ACTACGCCCTTGTGGAGTCCACTTCCACCCTGGGATTCGTCCAGTCTCCGAGT 1360  
Db 1435 ATTATGCTTCAATTAGTAATCAATCACTATGGGGTTTGTACCTGTAAAGTGAAT 1494  
QY 1361 AATAG 1365  
Db 1495 AAAAG 1499  
RESULT 15  
US-09-611-419A-11  
; Sequence 11, Application US/09611419A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Leonard A.  
; APPLICANT: Byrne, Michael P.  
; APPLICANT: Middlebrook, John L.  
; APPLICANT: Lapeotiere, Hugh  
; APPLICANT: Clayton, Michael A.  
; APPLICANT: Brown, Douglas R.  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; FILE REFERENCE: A33626 067252.0105  
; CURRENT APPLICATION NUMBER: US/09/611.419A  
; CURRENT FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: PCR/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/146,192  
; PRIOR FILING DATE: 1999-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1374  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: Synthetic construct based on BONTA Hc  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10)...(1362)  
US-09-611-419A-11  
Query Match 30.6%; Score 420.2; DB 23; Length 1374;  
Best Local Similarity 63.0%; Pred. No. 2.1e-78;  
Matches 754; Conservative 0; Mismatches 398; Indels 45; Gaps 5;  
QY 4 TTCAGATGACCATCCATTCACATCTTCTCCTACACCAACTCTTGTGAAGGAC 63  
Db 40 TTCGAGAACCATGTCATTCACATCTTCTCCTACACCAACTCTTGTGAAGGAC 99  
QY 64 ATCATCAAGGAGTACTTACACATCTACAGACTCCAGACTCTCCCTCCAGAACCGT 123  
Db 100 ATCATCAAGGAGTACTTACACATCTACAGACTCCAGACTCTCCCTCCAGAACAG 159  
QY 124 AAGAACACCTTGTGTGACACCTCCGGTTACAAACGCCGAGGTTCCGAGGAGGTCACGTC 183  
Db 160 AAGAACGCTTGTGTGACACCTCCGGTTACAAACGCCGAGGTCGAGTACGTCACGTC 219  
QY 184 CAGCTGAACCAATCTCCATTCAGCTTCCGTTGGTTCTCCGGTGAGGACAGAGGT 243

Db 220 CAGTTGAACACCATCTACACCAACGACTTCAAGTTGTCTCTTCCGGTGA-----C 270  
QY 244 AAGGTATCGTCAACCAAGAGAACATCGTCTACAACTCCATGTCAGGATCTCTCTCC 303  
Db 271 AGATCATCGTCAACTTGAACAAACACATCTGTACTCCGCCCTCTACGAGAACTCTCT 330  
QY 304 ATCTCTTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGC---CAGGTTACACCATC 360  
Db 331 GTCTCTTCTGATCAAGATCTCCAAAGACTTGACCAACTCCCAACAGAGTACACCATC 390  
QY 361 ATCGACTCCGTCAAGAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTCTGTC 420  
Db 391 ATCAACTCCATCGACGAGACTCCGGTTGGAAGTTGTATCCGTAACGGTAACATCGAG 450  
QY 421 TTACCCCTGAACAGAGAGGAGTCCGAGAGTCCATCAACTTCTCTTACGACATCTCC 480  
Db 451 TGGATCTTGCAGGAGTCAACCGTAAAGTACAAGTCTCTGATCTCTTCGACTTCTCCGAGTCC 510  
QY 481 AACACGCTCTCTGGTTA---CAACAAGTGGTTCTTCGTCACCGTCACCAACAACATGATG 537  
Db 511 TTGTCCACACCGGTTACACCAAGTGGTTCTTCGTCACCATCACCACAAACATCATG 570  
QY 538 GGTAAATGAAGATCTACATCAACGGTAAAGTGTGATCGACACCATCAAGGTCAAGGATG 597  
Db 571 GGTATCATGAAGTTGTATCATCAACGGTGAAGTGAAGCAGTCCCAAGAGATCGAGGACCTG 630  
QY 598 ACCGTTATCAACTTCTCCAAGACCATCACTTCGAGATCAACAAGATCCCGACACACCGGT 657  
Db 631 GACGAGTCAAGTGGACAGACCATCTCTTCGGTATCGAGGAGAACATCGA-----683  
QY 658 CTGATCACTCCGACTCCGACAAACATCAACATGTGTGATCCGTGACTTCTACATCTTCGCC 717  
Db 684 -----CGAGAACAGATGTTGTGATCCGTGACTTCAACATCTTCTCC 726  
QY 718 AAGGAGTTGGAGGTAAAGACATCAACATCTCTTCACTCTCTTCAGTACACCAACAGTC 777  
Db 727 AAGGAGTCTCCAAGGAGACATCAACATCTCTTACGAGGTCAGATCCCTGAGGAGAGTC 786  
QY 778 GTCAGGAGTACTGGGTAAACGACCTGAGATACACAAAGGAGTACTACTATGTTCAACATC 837  
Db 787 ATCAAGGAGTACTGGGTAAACCCACTGAAGTTGACACACCGAGTACTACTATCATCAAGAC 846  
QY 838 GACTACTTGAACAGATACATGTACGCCAACTCCACAGATCTCTTCAACACACAGAGT 897  
Db 847 AACTACATCGACCGTTACATCGCCCGAGAGTCCAAAGTCTCTGGTCCGATACCT 906  
QY 898 AACACACAGACTTCAACGAGGTTTACAAGATCATCATCAAGCGTATCAGAGGTAAACACC 957  
Db 907 GACCGTTCCAAGTGTACACCGGTAAACCTATCAACCATCAAGTCGGTCTCGGACAAAG 966  
QY 958 AAGCACACAGAGTCAAGGAGTGTGATCTCTGATCTTCGACATGACTATCAACAACAG 1017  
Db 967 CTTACTCCGATCTCTGAAACCGTGAACATCATCTCTGACATGCTGTACAACTCCCGT 1026  
QY 1018 GCTTACAACCTGTTTATGAAGAAGAGAGACCATGTACGCCGAGCAA-----CCACTCCACC 1071  
Db 1027 AGTATCATGATCATCGGTACACCGACACCATCTACGCCACCCAGGGTGGTACTGTCTCC 1086  
QY 1072 GAGGACATCTAGCCATCGGTCTGGTGTGAGAGACCAAGGACATCAAGCAGACATCATC 1131  
Db 1087 CAGAACTGTGTCTAGCCCTGAAGCTGACGTCCAACTGGGTAACTACGGTATCGGTATC 1146  
QY 1132 TTCACAGTCCGCCATGAACAACACTTACTACTACTCTCCAGATCTTCAAGTCC 1188  
Db 1147 TTCTCCATCAAGAACATCTCTTCCAAGAACAGTACTCTCCAGATCTTCTCTCTCC 1203

Search completed: November 7, 2002, 14:39:59  
Job time : 3190 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:26:05 ; Search time 34 Seconds  
(without alignments)  
4435.920 Million cell updates/sec

Title: US-09-910-186A-9  
Perfect score: 1371  
Sequence: 1 gaattacgagaccatccc.....tctccgagtaataagggaattc 1371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 193892 seqs, 55004114 residues

Total number of hits satisfying chosen parameters: 387784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*  
1: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	6.9	135259	6	US-10-240-425-1585
2	58.2	4.2	3028	5	US-09-548-938A-2
3	50.6	3.7	2835	6	US-10-092-411A-1515
C 4	50.6	3.7	3561	6	US-10-092-411A-1685
5	50.6	3.7	6414	6	US-10-092-411A-1685
6	50.4	3.7	15645	6	US-10-085-198-111
7	48.4	3.5	1452	6	US-10-270-223-1
C 8	47.6	3.5	783	6	US-10-092-411A-1670
9	46.8	3.4	8249	6	US-10-240-955-138
10	46	3.4	2466	6	US-10-267-255-83
11	46	3.4	18034	6	US-10-267-255-75
12	45.8	3.3	1092	6	US-10-270-223-5
13	45.6	3.3	1821	6	US-10-267-255-82
14	44.4	3.2	3401	6	US-10-230-437-123
15	44.4	3.2	3401	6	US-10-125-923A-411
16	43.2	3.2	8833	6	US-10-240-425-1410
17	42.8	3.1	2175	6	US-10-267-255-85
18	42.8	3.1	3878	6	US-10-264-237-907
19	42.8	3.1	4174	6	US-10-194-967-5
20	41.8	3.0	2334	6	US-10-131-813A-129
21	41.8	3.0	2334	6	US-10-131-819A-129
22	41.8	3.0	2334	6	US-10-131-823A-129
23	41.8	3.0	2334	6	US-10-131-824A-129
24	41.8	3.0	2334	6	US-10-131-826A-129
25	41.8	3.0	2334	6	US-10-131-829A-129
26	41.8	3.0	2334	6	US-10-125-926A-129

ALIGNMENTS

RESULT 1

US-10-240-425-1585  
; Sequence 1585, Application US/10240425  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Amanda

; APPLICANT: Boland, Joseph F.  
; APPLICANT: Lord, Reginald V.  
; APPLICANT: Alvarez, Chris  
; APPLICANT: Wetzel, Jon C.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Vockley, Joseph G.  
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue  
; FILE REFERENCE: 44521-5028  
; CURRENT APPLICATION NUMBER: US/10/240,425  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: PCT/US01/09847  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 60/193,446  
; PRIOR FILING DATE: 2000-03-31  
; NUMBER OF SEQ ID NOS: 1588  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1585  
; LENGTH: 135259  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. Z83838  
US-10-240-425-1585

Query Match 6.9%; Score 95; DB 6; Length 135259;

Best Local Similarity 46.4%; Pred. No. 7.6e-15;

Matches 391; Conservative 0; Mismatches 440; Indels 12; Gaps 2;

QY	249	CATCGTCAACCCAGAGAGACATGCTACACTCCATGTCACGAGTCCTTCTCCATCTC	308
DB	69406	CATCATCATCATCTACTGTGATCATTTACCACCATCACCACCATGATCATCATATATCAC	69465
QY	309	CTTCTGGATCAGAAACAACAAGTGGTCTCCAACTGTCAGGTTACACCATCATCGACTC	368
DB	69466	CACCGCCACCATCATCTGTGATCAGCACCACCATCACCACCATGATCATC	69525
QY	369	CGTCAAGACAACTCGGTGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC	428
DB	69526	CATTATCACCACCGCCACCATCATCTGTCACGAGTCATCATCATCATCATCATCAT	69585
QY	429	GAAGCAGACGAGGAGTCCGAGCAGTCATCATCACTCTCCCTACGACATCTCCAAACGC	488
DB	69586	TATCACCACCATCATCATCTGTCATCAGCACCACCATCATCATCATCATCATCATCAT	69645

```
OY 489 TCTGGTTACAAAGTGGTTCTTCGTACCGGTACCAACAAAGATGAGGGTAACATGAA 548
      |||||
Db 69646 CACCACGATACCAA-----CGTCACCAACCAACCAACCACTGTGATCAGCACCAT 69696
OY 549 GATCTACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAA 608
      |||||
Db 69697 CACCACCATCTACTACACCATCAACCAACCATGATCAACATATCACCACCAACCA 69756
OY 609 CTTCTCAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCCTGATCACCTC 668
      |||||
Db 69757 CATCACTGTGATCAGCACCATCAACCAACCATTAGCACCATCAACCAACCAACCA 69816
OY 669 CGACTCCGACATCAACATGTTGGTCCGTGACTTCTACATCTTCGCAAGGATTTGA 728
      |||||
Db 69817 CGTCACCAACCAACCATCACTGATGAGCACCATCAACCAACCATCAACCAACCA 69876
OY 729 CGGTAAAGACATCAACATCTCTTCAACTCTTTCAGTACCAACAGTGTGTCAGAGACTA 788
      |||||
Db 69877 CACCATTATCACCACCAACCACTACGGTGTGATGACCACTCACCACCATCAACCA 69936
OY 789 CTGG--GGTAACGACCTGATACACACAGGAGTACTACATGGTCAACATCGACTCTT 845
      |||||
Db 69937 CATGATCACCATCTCACCACCAACCATCACTGTGATCAGCACCATCATCACCACCA 69996
OY 846 GAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACCAAGGTAACAA 905
      |||||
Db 69997 CACCACCAACCATGATGCCATTATCACCATCACTGTGATCAGCACCATCAACCA 70056
OY 906 CGACTTCAACGAGGTTACAGATCATCATCAAGCGTATCAAGGTTAACAACCAAGCAC 965
      |||||
Db 70057 CACCATAATCACCACCAATATCACCATTATCATCACCACCATCAACCATCACTGTG 70116
OY 966 CAGAGTCAGAGGTTGATCACTCTGACTTCGACATGATATCAACAAGGCGCTCAAA 1025
      |||||
Db 70117 CACCATCACCATCATCACCATCATATTACAACCATCAACCATCAACCATCATCA 70176
OY 1026 CTTGTCATGAGACAGACCATGTAGCGGACACCACTCGACCGGAGGATCTAGCC 1085
      |||||
Db 70177 CACCACCAACCATCACTGTGATGATCAGCACCACCATCATTAACCATCAACCATCA 70236
OY 1086 CAT 1088
      |||
Db 70237 CAT 70239
```

## RESULT 2

```
US-09-548-938A-2
; Sequence 2, Application US/09548938A
; GENERAL INFORMATION:
; APPLICANT: EMALFARB, MARK AARON
; APPLICANT: BURLINGAME, RICHARD PAUL
; APPLICANT: OLSON, PHILIP TERRY
; APPLICANT: SINITSYN, ARKADY PANTELEIMONOVICH
; APPLICANT: PARRICHE, MARTINE
; APPLICANT: BOUSSON, JEAN CHRISTOPHE
; APPLICANT: PYNNONEN, CHRISTINE MARIE
; APPLICANT: PUNT, PETER JAN
; APPLICANT: VAN-ZEIJL, CORNELIA MARIA JOHANNA
; TITLE OF INVENTION: TRANSFORMATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI
; FILE REFERENCE: 3123-4001
; CURRENT FILING DATE: 2000-04-13
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3028
; TYPE: DNA
; ORGANISM: Chrysosporium lucknowense
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (46)..(47)
; OTHER INFORMATION: a, t, c, g, other or unknown
; FEATURE:
```

```
; NAME/KEY: modified_base
; LOCATION: (374)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-09-548-938A-2
```

Query Match 4.2%; Score 58.2; DB 5; Length 3028;

Best Local Similarity 44.3%; Pred. No. 4.4e-06;

Matches 292; Conservative 0; Mismatches 358; Indels 9; Gaps 1;

```
OY 364 GACTCGGTCTCAAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTTCTAGCATCTCCAA 423
      |||||
Db 1729 GCCAAGCGCAAGTCTATCCGCGCCACACCCCTCTCTGSCACTCTCAGTGGCGGAGTGG 1788
OY 424 ACCTTGAGCAGAACAGGACATCCGAGCAGTCCATCAACTTCTCTAGCATCTCCAA 483
      |||||
Db 1789 GTGAGAACATCAACACCGCAACACCTTGACCCAGGTATCGAGAACCCACCTACCA 1848
OY 484 AAGCTCTCTGGTTACAAAGTGTCTTCTCGTACCGTCCACCAACAAATGATGGGTAA 543
      |||||
Db 1849 CTTGTCACTCGTTACAAGGCAAGATCTCTCACTGGGAGCTGTTAAAGAGATCTTTGCC 1908
OY 544 ATGAGATCTACATCAACGGTAACTGATCGACACCATCAAGGTCAAGGATTTGACCGGT 603
      |||||
Db 1909 GAGGACGGCTCGCTCCGCGACACGCTTTCAGCCGGCTCTCGGGAGGAGCTTTGTGCGG 1968
OY 604 ATCAACTTCTCCAAAGACCATCACTTTCGAGATCAACAAGATCCACAGACACCGGTCTGATC 663
      |||||
Db 1969 ATGCGCTTCGCGCGCGCGCGCGCGCGGATCCCAAGCGCAAGCTCTACATCAACGACTAC 2028
OY 664 ACCTCGACATCCGACAAATCAACATGTGATCGGATCGGATCTTCTATCTTTGCCAAGGAG 723
      |||||
Db 2029 AACTCGACATTTCCCAACTACGCAAGGTGACCCGGGATGTGCGAGAGGTCAACAA 2088
OY 724 TTGACCGGTAAAGACATCAACATCTCTTCACTCTTCACTGATCAACCAACGCTCGTCAAG 783
      |||||
Db 2089 TGGATCCCGCAGGGCATCCGATCGACGGCATCGCACCCATGCGCCACTGCGCGGGCCC 2148
OY 784 GACTACTGGGGTA-----ACGACCTGAGATACAAAGAGTACTACATGTGTCAAC 834
      |||||
Db 2149 GCGGGTGGAAACACGCGCGCGCGCGCGCGCGCGCTCAAGGGCCCTCGCGCGGCGCAAC 2208
OY 835 ATCGACTACTTGAACAGATACATGTACGGCAACTCCAGACAGATCGTCTTCAACACCA 894
      |||||
Db 2209 GTCAGGAGATCGCATCACCAGCTCGACATCGCGGGCGCTCGCCACGACTACCTC 2268
OY 895 CGTAACAAACAGCTTCAACGAGGTTACAGATCATCATCAAGCGTATCAGAGTAA 954
      |||||
Db 2269 ACCGTATGAACGCTCGCTCCAGGTCTCCAAAGTGGTGGCATCACCCTCTGCGGCGCTC 2328
OY 955 ACCAAGACACACAGATGATCAGAGGTGGTGGATCTCTTCTGACATGACTATCAACA 1013
      |||||
Db 2329 TCTGACAAAGCAGCTGGAGGTGACAGCAGCAACCCGCTCTCTTTCGACAGCACTACCA 2387
```

## RESULT 3

```
US-10-092-411A-1515/c
; Sequence 1515, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: 032796-101
; CURRENT FILING DATE: 2002-03-07
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 1515
; LENGTH: 2835
```



; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1515

```
Query Match          3.7%; Score 50.6; DB 6; Length 2835;
Best Local Similarity 41.8%; Pred. No. 0.00041;
Matches 378; Conservative 0; Mismatches 524; Indels 3; Gaps 1;

QY 286 ATGTACGAGTCTCTCCATCTCTCTGATCAGATCAACAAAGTGGGTCTCCAACTTG 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2654 ATCAGCGAGTACGTCGACATCTGATGTGATCAGAGTCAGCAAGAAAGAGTGCAC 2595

QY 346 CCAGGTATACCATCATCGACTCGGTCAAGAAACAATCCGGTGGTTCACGCTATCATC 405
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2594 GAGTAAAGTAGTCGTCAATTAAGTGAATCGAAGCAAGAGCGTTTCAGATTCAGCAAGCGT 2535

QY 406 TCCAACTTCCTGGTCTTCCACCTGAAGCAAGAGGAGCTCCGAGCAGTCCATCAACTTC 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2534 GTCACGTCAGAAAGTGGTCAACATCAACAAAGTGTGAGTGGTCAACAGTACAGCAT 2475

QY 466 TCCACGACATCTCCAAACAGCTCCCTGGTTTAAACAAGTGGTCTTCGTCCCGCTCAC 525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2474 TTCAGACTCAACGAGCAGATCGAGCTGTGATAGCGCATCAATAAAGCGAGGGAATCAGC 2415

QY 526 AACACATGATGGGTACATGAGATACATCAACGTTAAGTGTGATCGACCATCAAG 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2414 AAGCACAAGTAAATATATACAAATCAGTGTGATCGTCAATTAAGTAAATCGCAGAGCAC 2295

QY 586 GTCAAGGAGTTGACCGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAGATC 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2354 GTCAAGGAGCGTAAAGTACCTCAATATCGCAAGTACGTCTATTAAGTAAATCGCAGAGCAC 2295

QY 646 CCAGACACCGGTCTGATCACTCCGACTCCGACCAACATCAACATGTGGATCCGTGACTTC 705
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2294 AAGCGTTTCAGATTCAACAAGTACATCGACATCCGACAGTGGCTCCACATCAACGAGTGA 2235

QY 706 TACATCTTCGCCAAGGAGTGTGGAGCTGAAGCATCAACATCTGTTCAACTCCTTTCGAG 765
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2234 GAGTGAGTCAGACAGTGCAGTACGTCTCATTAAGTGAATCGACAAGTACAAGCGTTTCAGA 2175

QY 766 TACACCAAGCTGTCAAGGACTACTGGGTAACAGACCTTGATGATCAACAGAGTACTAC 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2174 TTCAACGAGTACGTGACATCCGACATCGATGATCAATGTCTAGCGAGGAGAGTGAATCAAA 2115

QY 826 ATGGTCAACATCGACTACTTGAACAGATACATGTACGCCAATCCAGACAGATCGCTTC 885
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2114 CAGTAAAGTAGTACGATTAAGTGAATCGACAAGTACGAGTCTTTCAGGCTCAACGAGTGC 2055

QY 886 AACACCGACGTAACAACAGCTTCAACGAGGTTTCAAGATCATCATCAAGCGTATC 945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2054 GTCGACATCCGACAGTGCATCAAGTCAACGAGTGAAGTGAAGTACATCGTTGAGTGA 1995

QY 946 AGAGTAAACCAACGACACACAGAGTGCAGAGTGGTGACATCTCTGATCTTCGACATGACT 1005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1994 GTCATTAAAGTGAATCGACAAGTACGAGTCTTTCAGGCTCAACGAGTGGTGCNATCCGA 1935

QY 1006 ATCAACAAACAGGCGCTCAACCTGTTTCATGAAGAACGAGACCATGTACCGCCGACACAC 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1934 CAGTGCATCAACGTCACAAACGAGAGTGAAGTCAACAAAGTGAAGTACATCGTTGAGTGA 1875

QY 1066 TCCACCGAGGACATCTACGCCATCGTCTCGGTGACGACACCAAGGACATCAACGACAC 1125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1874 ATCATTAAGACACA---AGCGTTTCAGATTCACAAGCGCGGTCAACGTCAGAAAGTGCATC 1818

QY 1126 ATCATCTTCCAGATCCAGCCAAATGAACAACACTTACTTACTTACGCTTCCGAGATCTTCAAG 1185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1817 CACATCAACGAGTGAGAGTGAGTCAACACAGTGCACGACATCGTTAAGTGGATCATTAAG 1758

QY 1186 TCCAA 1190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1757 TACAA 1753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 4

US-10-092-411A-1685/c  
; Sequence 1685; Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 1685  
; LENGTH: 3561  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1685

## Query Match 3.7%; Score 50.6; DB 6; Length 3561;

Best Local Similarity 41.8%; Pred. No. 0.00046;  
Matches 378; Conservative 0; Mismatches 524; Indels 3; Gaps 1;

```
QY 286 ATGTACGAGTCTCTCCATCTCTCTGATCAGATCAACAAAGTGGGTCTCCAACTTG 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3450 ATCAGCGAGTACGTCGACATCTGATGTGATCAGAGTCAGCAAGAAAGAGTGCAC 3391

QY 346 CCAGGTATACCATCATCGACTCGGTCAAGAAACAATCCGGTGGTTCACGCTATCATC 405
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3390 GAGTAAAGTAGTCGTCAATTAAGTGAATCGAAGCAAGAGTGTGATCGAGTCAAC 3331

QY 406 TCCAACTTCCTGGTCTTCCACCTGAAGCAAGAGGAGCTCCGAGCAGTCCATCAACTTC 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3330 GTCAACGTCAGAAAGTGGTCAACATCAACAAAGTGTGAGTGGTCAACAGTACAGCAT 3271

QY 466 TCCATACGACATCTCCAAACAGCTCCCTGGTTTAAACAAGTGGTCTTCGTCCCGCTCAC 525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3270 TTCAGATCAACGAGCAGATCGAGCTGTGATAGCGCATCAATAAAGCGAGGATCAGC 3211

QY 526 AACACATGATGGGTACATGAAGATCTACATCAACGAGTGAAGTGTATCGACACCATCAAG 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3210 AAGCACAAGTAAATATATACAGATCAGTACGTACGTCTGATAGCGCATCAAC 3151

QY 586 GTCAAGGAGTGTACCGGTATCACTCTTCCAGACCATCACTTCGAGATCAACAGATC 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3150 GTCAACGAGCGTAAAGTACCTCCAAATAGCGCAAGTACGTCTTAAGTAAATCGCAAGTGC 3091

QY 646 CCAGACACCGGTCTGTGATCACTCCGACTCCGACCAACATCAACATGTGGATCGGTGACTTC 705
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3090 AAGCGTTTCAGATTCAACAAGTACATCGACATCCGACAGTGGCTCCACATCAACGAGTGA 3031

QY 706 TACATCTTCGCCAAGGAGTGGAGGTGAGGATCAACATCAACATCTCTGTTCACTTCCCTTTCAG 765
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3030 GAGTGAGTCAAGAGTGCAGTACGCTCATTAAGTGAATCGCAAGTACAAAGCGTTTCAGA 2971

QY 766 TACACCAACGCTCGTCAAGGACTACTTGGGGTAACGAGCTGAGATACAAAGGAGTACTAC 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2970 TTCACAGGATACGTCGACATCCGACAGTGCATCAATGTACGAGCGAGAGTGAATCAAA 2911

QY 826 ATGTCAACATCGACTACTTGAACAGATACATGTAGCGCACTCCAGACAGATCGCTTTC 885
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2910 CAGTAAAGTAGTCGTCAATTAAGTGAATCGAAGTACGAGTCTTTCAGGCTCAACGAGTGC 2851

QY 886 AACACCGAGGATCAACAACAGCTTCAACGAGGTTTAAAGATCATCATCAAGCGTATC 945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2850 GTCGACATCCGACAGTGCATCAACGTCACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2791

QY 946 AGAGGTAAACCAACGACACACGAGTGGTGGATCGAGTGTGAGTGTGAGTGTGAGTGTG 1005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



Db 2790 GTCATTAAAGTGAATCGACAAAGTACGAGTCTTTTCAGGCTCAACGAGTGCCTCGACATCCGA 2731  
 QY 1006 ATCAACAAAGGCGCTTACAACCTGTTTCATGAAGAAGAGAGACCATGTACCGCCGCAACCCAC 1065  
 Db 2730 CAGTGCATCAACGAGGAGTACTCAACAAAGTGAAGTACATCGTTGATGA 2671  
 QY 1056 TCCACGAGGACATCTACCGCATCGGTCGCTGAGAGACACCAAGACATCAACGACAAAC 1125  
 Db 2670 ATCATTAAAGCACA---AGCGTTTCAGATTCAACAAAGCGGTCAACGTCAGAAGTGCATC 2614  
 QY 1126 ATCATCTCCAGATCCAGCAACTGAACAACTTACTACTACGCTTCCAGATCTTCAAG 1185  
 Db 2613 CACATCAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2554  
 QY 1186 TCCAA 1190  
 Db 2553 TACAA 2549  
 RESULT 5  
 US-10-092-411A-1626  
 ; Sequence 1626, Application US/10092411A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 032796-101  
 ; CURRENT APPLICATION NUMBER: US/10/092,411A  
 ; CURRENT FILING DATE: 2002-03-07  
 ; PRIOR APPLICATION NUMBER: US 09/134,001  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5676  
 ; SEQ ID NO 1626  
 ; LENGTH: 6414  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis  
 US-10-092-411A-1626  
 Query Match 3.7%; Score 50.6; DB 6; Length 6414;  
 Best Local Similarity 41.8%; Pred. No. 0.00062;  
 Matches 378; Conservative 0; Mismatches 524; Indels 3; Gaps 1;  
 QY 286 ATGACGAGTCCCTTCCATCTCTCTGATCAGATCAACAAAGTGGGTCCCAACTTG 345  
 Db 2595 ATCAGCGAGTACGTCGACATCTGATAGTGCATCAGATCAGCAAGAAAGAGTGAAGTCAAC 2654  
 QY 346 CCAGGTACACCATCATCGACTCGGTCAAGAACTCCGGTTGGTCCATCGGTATCATC 405  
 Db 2655 GAGTAAAGTACGTCATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2714  
 QY 406 TCCAACTCTCTGCTTCCACCTCAAGCAAGAGAGACTCCGAGAGTCCCACTCACTTC 465  
 Db 2715 GTCAAGGTCAAGAAAGTGGTCAACATCAACAAAGTGTGAGTGGTCAACAAAGTACAGAT 2774  
 QY 466 TCCAGGACATCTCCACAGAGCTCTGGTTTACAAAGTGGTGTTCGTACCGTCAAC 525  
 Db 2775 TCCAGGACATCAAGAGCAGATCGAGTGTGATAGCGCATCAATAAAGCGGAGCAATCAGC 2834  
 QY 526 AACACATGATGGTGAATGAAGATGATATCAACGCTAAGTGTATCGACACCATCAAG 585  
 Db 2835 AAGCACAAGTAACTATATCAAGTACAGTACGTCGACGCTGTATAGCGCATCAAC 2894  
 QY 586 GTCAGGAGTGGACCGGTATCACTTCCAGACCATCACTTCGAGATCAACAAAGATC 645  
 Db 2895 GTCAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2954  
 QY 646 CCAGACACCGGTGTGATCACTCGGATCCGACAAACATCAACATCACTCGGTGACTTC 705  
 Db 2955 AAGGTTTCAAGTCAACAAAGTACATCCGACATCCGACATCCGACATCCGACATCCGAC 3014

QY 706 TACATCTTCCCAAGAGGTTGGACGTTAAGGACATCAACATCTCTTCAACCTCCCTTCCAG 765  
 Db 3015 GAGTGTAGTACAGAGTCAAGTACGTCATTAAGTGAATCGCAAGTACAAGGTTTTCAGA 3074  
 QY 766 TACACCAACGCTCTCAGGACTACTGGGTAAAGGCTTACGACCTGAGATCAACAAAGGAGTACTAC 825  
 Db 3075 TTAACGAGTACGTCGACATCCGACAGTGCATCAATGTCCAGCGGAGAGTGAATCAAA 3134  
 QY 826 ATGCTCAACATCGACTACTTGAACAGATACATGTAGCCCAACTCCAGACAGATCGTCTTC 885  
 Db 3135 CAGTAAAGTACGTCATTAAGTGAATCGCAAGTACGAGTCTTTCAGGCTCAACGAGTGC 3194  
 QY 886 AACACGAGAGTACAAACACAGACTTCAACGAGGTTTCAAGATCATCATCAAGGATAC 945  
 Db 3195 GTCGACATCCGACAGTGCATCAACGTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 3254  
 QY 946 AGAGTAAACACCAACGACACAGAGTGCAGAGTGGTGAACATCTCTTACGATGACT 1005  
 Db 3255 GTCATTAAAGTGAATCGACAGTACGAGTCTTTCAGGCTCAACGAGTGGTGCAGATCCGA 3314  
 QY 1006 ATCAACAAAGGCTTACAACTGTTTCATGAAGAGAGGACCATGTACGCGCAACACCAC 1065  
 Db 3315 CAGTGCATCAACGCTCAACAAAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3374  
 QY 1066 TCCACCGAGGACATCTACGCGCTCTGCGTGAAGCAAGGAGATCAACAAAGGAGTCAAC 1125  
 Db 3375 ATCATTAAAGCACA---AGCGTTTCAAGTTCACAAAGCGGTCAACGAGTGCATC 3431  
 QY 1126 ATCATCTCCAGATCCAGCAACTGAACAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1185  
 Db 3432 CACATCAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3491  
 QY 1186 TCCAA 1190  
 Db 3492 TACAA 3496

RESULT 6  
 US-10-085-198-111  
 ; Sequence 111, Application US/10085198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alsobrook et al  
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 21402-279  
 ; CURRENT APPLICATION NUMBER: US/10/085,198  
 ; CURRENT FILING DATE: 2002-02-25  
 ; PRIOR APPLICATION NUMBER: 60/271,646  
 ; PRIOR FILING DATE: 2001-02-26  
 ; PRIOR APPLICATION NUMBER: 60/276,401  
 ; PRIOR FILING DATE: 2001-03-16  
 ; PRIOR APPLICATION NUMBER: 60/311,981  
 ; PRIOR FILING DATE: 2001-08-13  
 ; PRIOR APPLICATION NUMBER: 60/312,858  
 ; PRIOR FILING DATE: 2001-08-16  
 ; PRIOR APPLICATION NUMBER: 60/271,840  
 ; PRIOR FILING DATE: 2001-02-27  
 ; PRIOR APPLICATION NUMBER: 60/277,324  
 ; PRIOR FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: 60/286,096  
 ; PRIOR FILING DATE: 2001-04-21  
 ; PRIOR APPLICATION NUMBER: 60/299,595  
 ; PRIOR FILING DATE: 2001-06-20  
 ; PRIOR APPLICATION NUMBER: 60/315,614  
 ; PRIOR FILING DATE: 2001-08-29  
 ; PRIOR APPLICATION NUMBER: 60/272,405  
 ; PRIOR FILING DATE: 2001-02-28  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 653  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 111  
 ; LENGTH: 15645  
 ; TYPE: DNA





```
Query Match 3.4%; Score 46; DB 6; Length 2466;
Best Local Similarity 43.2%; Pred. No. 0.0061;
Matches 269; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

Qy 470 ACAGATCTCCACACAGCGTCTCGGTGTACACAAAGTGGTCTTCGTACACCGTCCACCAACA 529
Db 1517 ACAGCAGCGCGACATCGCGGTCTGTACAGACATGGTTCGCACTGAGGACGCGCCCAACC 1576
Qy 530 ACATGATGGTGAACATGAAGATCTACATCAACGGTAAAGTGTATCGACACCATCAAGGTCA 589
Db 1577 GCACCAAGTTGTACGAGTTTACCAGCAGCCGAGCGAGGATTCACAGCCGGTCAAGGTCT 1636
Qy 590 AGAGTTGACCGGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAGATCCAG 649
Db 1637 GGGACACAGCAGCAGCCCGCTCAAGAGCTGGAATGGCCCTCCAGCAAGCTGACCGTGG 1696
Qy 650 ACACCGGTCTGATCACTCCGACTCCGACATCAACATCAACATGATGGATCGGTACTTCTACA 709
Db 1697 GCGACTTCGAGCGGAGCGGCAAGCGCGACATCGCGTCTGTACGACATACGCGAGGACG 1756
Qy 710 TCTTCGCCAAGGAGTTGGAGGTTAAGAGATCAACATCACTTCTGTACGACTACGCGAGTACA 769
Db 1757 GCGACCGCAGCCGTACGCGGCTGTGACCTTCCAGCAGCCGCTGCGGCTTCCAGCGCC 1816
Qy 770 CCAAGCTCTCAAGGACTACTGGGTAAAGCAGCTGAGATACACAGGAGTACTACATGG 829
Db 1817 CCAAGCTGTGTGGACAGCAAC---AAGACCCGGTCAAGAGTGGAGTGGAGACGCA 1873
Qy 830 TCACATCGACTACTTGAACAGATATGATGACCCAACTCCAGACAGATCGCTTTCACA 889
Db 1874 GCAAGCCACCGTCGGGACTTCAAGGGGAGCGCAAGCGGACATCGCGTCTCTACG 1933
Qy 890 CCAGACGTAAACAACAGACTTCAAGGAGGTTACAGATATCATCAAGCGTATCAGAG 949
Db 1934 ACATGGTTCGACGAGGAGCGCGCAACCGCACCAAGTGTTCACCTTACCGCGCAGG 1993
Qy 950 GTACACCAACAGCACACAGATGAGAGTGGTGCATCTCTGATTCGACATGACTATCA 1009
Db 1994 CGACCGGTTTCAACAGCCCGTCAAGTGTGGGACAGCAACAGCAGCCCGTGAAGAGCT 2053
Qy 1010 ACACAGGCGCTACACCTCTTCAAGAGAGGTTACAGATATCATCAAGCGTATCAGAG 1069
Db 2054 GGAATCGGAGCGCGTCAAGTGTGGAGGCGACTTCAACGGCGAGCGCAAGCGGACA 2113
Qy 1070 CCGAGGACATCTAGCCATCGG 1091
Db 2114 TCGGGGTGTGTAGACTACGG 2135

RESULT 11
US-10-267-255-75
; Sequence 75, Application US/10267255
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 75
```

```
; LENGTH: 18034
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (302)...(302)
; OTHER INFORMATION: n is a or t or g or c
US-10-267-255-75

Query Match 3.4%; Score 46; DB 6; Length 18034;
Best Local Similarity 43.2%; Pred. No. 0.017;
Matches 269; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

Qy 470 ACAGATCTCCACACAGCGTCTCGGTGTACACAAAGTGGTCTTCGTACACCGTCCACCAACA 529
Db 13031 ACGGCAGCGCGACATCGCGGTCTGTACGACATGGTTCGCACTGAGGACGCGCCCAACC 13090
Qy 530 ACATGATGGTGAACATGAAGATCTACATCAACGGTAAAGTGTATCGACACCATCAAGGTCA 589
Db 13091 GCACCAAGTTGTACGAGTTTACCAGCAGCCGAGCGGATTCACAGCCGGTCAAGGTCT 13150
Qy 590 AGAGTTGACCGGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAGATCCAG 649
Db 13151 GGGACACAGCAGCAGCCCGCTCAAGAGCTGGAATGGCCCTCCAGCAAGCTGACCGTGG 13210
Qy 650 ACACCGGTCTGATCACTCCGACTCCGACATCAACATCAACATGATGGATCGGTACTTCTACA 709
Db 13211 GCGACTTCGAGCGGAGCGGCAAGCGCGACATCGCGTCTGTACGACTACGCGAGGACG 13270
Qy 710 TCTTCGCCAAGGAGTTGGAGGTTAAGAGATCAACATCACTTCTGTACACTCCCTTTCAGTACA 769
Db 13271 GCGACCGCAGCCGTACGCGGCTGTGACCTTCCAGCAGCCGCTGCGGCTTTCACCGGCC 13330
Qy 770 CCAGCTCGTCAAGGACTACTGGGTAAAGCAGCTGAGATACACAGGAGTACTACATGG 829
Db 13331 CCAAGCTGGTGTGGAGAGCAAC---AAGACCCGGTCAAGAGTGGAGTGGAGACGCA 13387
Qy 830 TCACATCGACTACTTGAACAGATATGATGACCCAACTCCAGACAGATCGCTTTCACA 889
Db 13388 GCAAGCCACCGTCGGGACTTCAAGGGGAGCGCAAGCGGACATCGCGTCTCTACG 13447
Qy 890 CCAGACGTAAACAACAGACTTCAAGGAGGTTACAGATATCATCAAGCGTATCAGAG 949
Db 13448 ACATGGTTCGACGAGGAGCGCGCAACCGCACCAAGTGTTCACCTTACCGCGCAGG 13507
Qy 950 GTACACCAACAGCACACAGATGAGAGTGGTGCATCTCTGATTCGACATGACTATCA 1009
Db 13508 CGACCGGTTTCAACAGCCCGTCAAGTGTGGGACAGCAACAGCAGCCCGTGAAGAGCT 13567
Qy 1010 ACACAGGCGCTACACCTCTTCAAGAGAGGAGTACAGATATCATCAAGCGTATCAGAG 1069
Db 13568 GGAATCGGAGCGCGTCAAGGTGTGGAGGCGACTTCAACGGCGAGCGCAAGCGGACA 13627
Qy 1070 CCGAGGACATCTAGCCATCGG 1091
Db 13628 TCGGGGTGTGTAGACTACGG 13649

RESULT 12
US-10-270-223-5
; Sequence 5, Application US/10270223
; GENERAL INFORMATION:
; APPLICANT: Biocimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR CON
; TITLE OF INVENTION: INTACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMA-TIO
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1092
```

```
; TYPE: DNA
; ORGANISM: Aequoria Victoria and Human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1092)
; OTHER INFORMATION:
US-10-270-223-5

Query Match      3.3%; Score 45.8; DB 6; Length 1092;
Best Local Similarity 47.4%; Pred. No. 0.0045;
Matches 137; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 678 CAACATCAACATGTTGGATCCGTGACCTTACATCTTCGCCAAGAGGTTGGACGGTAAGA 737
Db 291 CACCATCTTCTTCAAGGACGACGCGCACTACAGACCCGCGGAGGTGAAGTTCGAGGG 350
QY 738 CATCAACATCCTGTTCAACTCTTTCAGTACACCAAGCTGTCGAAGACTTACTGGGTA 797
Db 351 CGACACCCCTGTTGAACCGCATCGAGCTGAAGGCGATCGACTTCAAGGAGGACGGCAACAT 410
QY 798 CGACCTGAGATACACACAGGAGTACTACATGGTCAACATCGACTACTTGNACAGATACAT 857
Db 411 COTGGGGCACAAGTGGAGTACAACTCAACAGCCCAACAGCTATATATCATGCGCGACAA 470
QY 858 STACCCCAACTCCAGACAGATCGCTTTCACACCCAGACAGTAAACAACAACGACTTCAACGA 917
Db 471 GCAGAGAAGCGCATCAAGGTTGAATTCAGATCCGCCCAACATCGAGGACGGCAGCGT 530
QY 918 GGGTTACAGATCATCATCAGCGTNTCAGAGGTAAACCAACCAAGCAC 966
Db 531 GCAGCTCCGCGACCACTACCAGCAGAACACCCCATCGGCGAGCGGCC 579

RESULT 13
US-10-267-255-82
; Sequence 82, Application US/10267255
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varcglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456Usl
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 82
; LENGTH: 1821
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-82

Query Match      3.3%; Score 45.6; DB 6; Length 1821;
Best Local Similarity 41.9%; Pred. No. 0.0066;
Matches 273; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

QY 440 AGGATCCGACGAGTCCATCACTTCTCCTACGACATCTCCAAACAGCTCTCGGTACA 499
Db 1031 AGCCAGGAAGCTCACTCCGCGCATCAACGCGGACGCAAGCGCGCATCGGGTGC 1090
QY 500 ACAAGTGTCTTCTGCTACCGCTCACCACCAACATGTTGGTAAACATCAATCATCA 559
Db 1091 TGTAGACTACGGCAAGGACGACGACCCACACCGGACAGGGCTCTGGACGTTACCAGCA 1150

; TYPE: DNA
; ORGANISM: Aequoria Victoria and Human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1092)
; OTHER INFORMATION:
US-10-270-223-5

Query Match      3.3%; Score 45.8; DB 6; Length 1092;
Best Local Similarity 47.4%; Pred. No. 0.0045;
Matches 137; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 678 CAACATCAACATGTTGGATCCGTGACCTTACATCTTCGCCAAGAGGTTGGACGGTAAGA 737
Db 291 CACCATCTTCTTCAAGGACGACGCGCACTACAGACCCGCGGAGGTGAAGTTCGAGGG 350
QY 738 CATCAACATCCTGTTCAACTCTTTCAGTACACCAAGCTGTCGAAGACTTACTGGGTA 797
Db 351 CGACACCCCTGTTGAACCGCATCGAGCTGAAGGCGATCGACTTCAAGGAGGACGGCAACAT 410
QY 798 CGACCTGAGATACACACAGGAGTACTACATGGTCAACATCGACTACTTGNACAGATACAT 857
Db 411 COTGGGGCACAAGTGGAGTACAACTCAACAGCCCAACAGCTATATATCATGCGCGACAA 470
QY 858 STACCCCAACTCCAGACAGATCGCTTTCACACCCAGACAGTAAACAACAACGACTTCAACGA 917
Db 471 GCAGAGAAGCGCATCAAGGTTGAATTCAGATCCGCCCAACATCGAGGACGGCAGCGT 530
QY 918 GGGTTACAGATCATCATCAGCGTNTCAGAGGTAAACCAACCAAGCAC 966
Db 531 GCAGCTCCGCGACCACTACCAGCAGAACACCCCATCGGCGAGCGGCC 579

RESULT 14
US-10-230-437-123
; Sequence 123, Application US/10230437
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3530P1C94
; CURRENT APPLICATION NUMBER: US/10/230,437
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
```

```
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 246
; SEQ ID NO 123
; LENGTH: 3401
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-230-437-123

Query Match          3.2%; Score 44.4; DB 6; Length 3401;
Best Local Similarity 44.3%; Pred. No. 0.019;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 730 GGTAAGGACATCAACATCTGTTCAACTCTTCCAGTACACCAAGCTGCTCAAGACTAC 789
Db 983 GGGTGCACCTGCGAAGCTGTCATCAACATGAGGGCACCACCAAGCTATCTCTCAAC 1042
QY 790 TGGGGTAACGACCTGAGATACACAAAGAGTACTATCATGTTCAACATCGACTTGAAC 849
Db 1043 AGCCTCAAGAAGATGCGAAGCTGACTGAGCTGAGCTGATCCGCTGGAGCGC 1102
QY 850 AGATACATGATGCCCACTCCAGACAGATCGTCTTCAACACAGAGCTTAACACAGCAC 909
Db 1103 ATCCCCCACTCCATCTGAGCTCCCAACCTCGAGAGATTGACCTCAAGGACAAAC 1162
QY 910 TTCAAGGAGGTTTACAGATCATCAAGCGTATCAGAGGTAACACCAACGACACACCA 969
Db 1163 CTCAAGACCATCGAGGAGATCATCAGCTTCCAGACCTGACCCGCTTAAAG 1222
QY 970 GTGAGAGTGGTGACATCTGTTACTTTCGACATGACTATCAACAAAGCCCTACACCTG 1029
Db 1223 CTGTGGTACAAACCATCGCTTACATCCCATCCAGATCGGCACTTCCCAACCTGGAG 1282
QY 1030 TTATGAAGAACGAGACCATGTACGCCGACAAACCACTCCACGAGGACATCTACGCCATC 1089
Db 1283 CGCTCTACCTGAACCGCAACAGATCGAGAGATCCCAACCTCTTCTACTGCCGC 1342
QY 1090 GGTCTGGTGAGCAGACCAAGGACATCAACGACACACATCATCTTCC 1135
Db 1343 AAGCTGGCTTACCTGGACCTCAGCCCAACAAACCTGACCTTCTCTCC 1388

RESULT 15
US-10-125-923a-411
; Sequence 411, Application US/10125923A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC79
; CURRENT APPLICATION NUMBER: US/10/125,923A
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
```

```
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 411
; LENGTH: 3401
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-125-923a-411

Query Match          3.2%; Score 44.4; DB 6; Length 3401;
Best Local Similarity 44.3%; Pred. No. 0.019;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 730 GGTAAGGACATCAACATCTGTTCAACTCTTCCAGTACACCAAGCTGCTCAAGACTAC 789
Db 983 GGGTGCACCTGCGAAGCTGTCATCAACATGAGGGCACCACCAAGCTATCTCTCAAC 1042
QY 790 TGGGGTAACGACCTGAGATACACAAAGAGTACTATCATGTTCAACATCGACTTGAAC 849
Db 1043 AGCCTCAAGAAGATGCGAAGCTGACTGAGCTGAGCTGATCCGCTGGAGCGC 1102
QY 850 AGATACATGATGCCCACTCCAGACAGATCGTCTTCAACACAGAGCTTAACACAGCAC 909
Db 1103 ATCCCCCACTCCATCTGAGCTCCCAACCTCGAGAGATTGACCTCAAGGACAAAC 1162
QY 910 TTCAAGGAGGTTTACAGATCATCAAGCGTATCAGAGGTAACACCAACGACACACCA 969
Db 1163 CTCAAGACCATCGAGGAGATCATCAGCTTCCAGACCTGACCCGCTTAAAG 1222
QY 970 GTGAGAGTGGTGACATCTGTTACTTTCGACATGACTATCAACAAAGCCCTACACCTG 1029
Db 1223 CTGTGGTACAAACCATCGCTTACATCCCATCCAGATCGGCACTTCCCAACCTGGAG 1282
QY 1030 TTATGAAGAACGAGACCATGTACGCCGACAAACCACTCCACGAGGACATCTACGCCATC 1089
Db 1283 CGCTCTACCTGAACCGCAACAGATCGAGAGATCCCAACCTCTTCTACTGCCGC 1342
QY 1090 GGTCTGGTGAGCAGACCAAGGACATCAACGACACACATCATCTTCC 1135
Db 1343 AAGCTGGCTTACCTGGACCTCAGCCCAACAAACCTGACCTTCTCTCC 1388

Search completed: November 7, 2002, 14:42:56
Job time : 189 secs
```



Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA sequencing by: Washington University Genome Sequencing Center Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov

Seq primer: T3 Et from Amersham  
High quality sequence stop: 527.

## FEATURES

source

```

1. .641
Location/Qualifiers
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Sugano SJD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site 1: DraIII (CACATGTG); Site 2: DraIII (CACATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGCGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGG, 3' site CACATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTAAAGCTGGG and 3' end primer CGACCTGACCTGAGGACA."
218 a 234 c 37 g 152 t
CGACCTGACCTGAGGACA."

```

BASE COUNT

ORIGIN

```

Query Match 7.8%; Score 107.2; DB 13; Length 641;
Best Local Similarity 49.7%; Pred. No. 3.9e-13;
Matches 308; Conservative 0; Mismatches 303; Indels 9; Gaps 1;

QY 270 CAGCTGTCACAACTCCATGTACGAGTCTTCCATCTCTCTGATCAGATCAACAA 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30 CATCATCTCTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTA 89

QY 330 GTGGGTCTCCAACTTCCAGGTTACACCATCATCGCTCGTCAAGAACAACTCCGGTTG 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 CTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTT 149

QY 390 GTCCATGGGTATCATCTCCAACTTCTGCTGCTTCCACCTGAGCAGACGAGGACTCGGA 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 CTTCCAAACAAACAACTTCTCCAGTATCATCACTTCTCCAAACAAACAACTTCTCCAA 209

QY 450 CGAGTCCATCAACTTCTCTAGGACATCTCCAAACAGCTCTGTTTACACAGTGGTT 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 CAACAACTTCTCCAGTATCATCACTTCTCCAAACAACTTCTCCAAACAACTTCTC 260

QY 510 CTTCTCCAGTCCATCAACATGATGGTGAACATGATGATGATGATGATGATGATGATG 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 CAGCTATCATCACTTCAACAAACAACTTCTCCAGTATCACTTCTCCAGCAGACACAT 320

QY 570 GATCCAGACCATCAAGGTCAGGAGTTCACCGGTATCACTTCTCCAGGACCATCACTT 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 CAACCTTCTCCAGTATCATCACTTCTCCAAACAAACAACTTCTCCAGTATCACTT 380

QY 630 CGAGATCAACAGATTCACAGACACCGGCTCATCACTTCCGACTCCGACATCAACAT 689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 CTTCCAAACAAACAACTTCTCCAGTATCACTTCTCCAAACAAACAACTTCTCCAG 440

QY 690 GTGGATCCGTGATCTTCACTTCTCGGCAAGAGTGTGAGCGGTGAAGACATCACTTCT 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 CTATCACTTCTCAACAAACAAACAACTTCTCCAGTATCACTTCTCAACAAACAA 500

QY 750 GTTCACTCTCTGAGTACACCAAGCTGCTCAAGACTACTGGGTGAAGACCTGAGATA 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 CTTCTCCAGTATCACTTCTCCAGCAGACAACTCACTTCTCCAGTATCACTTCTCA 560

QY 810 CAACAGGAGTACTATGTTGTTCAACATCGACTTCTTGAACAGATATGATGACGCCAACTC 869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 CAACAAACAACTTCTCCAGTATCACTTCAACAAACAAACAACTTCTCCAGCTA 620

QY 870 CAGACAGATCGCTTCAACA 889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 621 CAACTTCTCCAGCAACAACA 640

RESULT 2
BH153606 906 bp DNA linear GSS 24-SEP-2001
LOCUS ENT8583TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION BH153606
VERSION BH153606.1 GI:15725323
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (Bases 1 to 906)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 15
High quality sequence stop: 733.
Location/Qualifiers
1. .906
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pROSL; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) 'Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 149 a 293 c 66 g 398 t
ORIGIN

Query Match 7.8%; Score 107; DB 17; Length 906;
Best Local Similarity 46.4%; Pred. No. 5e-13;
Matches 350; Conservative 0; Mismatches 405; Indels 0; Gaps 0;

QY 247 GTCATGTCACCCAGAGAGACATCGTTACAACTCCATGTACGAGTCTTCCATC 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 GTCCTTCTCATCGCTTCTTTCATCGTCTTCTTTCATCGTCTTCTTTCATC 100

QY 307 TCCCTTCGATCAGAATCAACAAGTGGTCTCCAACTTGCAGGTTACACCATCATCGAC 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 GTCCTTCTCATCGTCTTCTTTCATCGTCTTCTTTCATCGTCTTCTTTCATC 160

QY 367 TCCGTCAGAGAACTCCGGTTGGTCCATCGTATCATCTCCAACTTCTCGTCTTCC 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





RESULT 4	BH146886	890 bp	DNA	linear	GSS 27-AUG-2001
LOCUS	ENTPK48TF	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
DEFINITION	genomic, DNA sequence.				
ACCESSION	BH146886	1	GI:15302963		
VERSION	BH146886.1				
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				
ORGANISM	Entamoeba histolytica				
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.				
AUTHORS	1 (bases 1 to 890)				
TITLE	Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.				
JOURNAL	Determination of clone end sequences from Entamoeba histolytica				
COMMENT	HMI:IMSS sheared DNA library (2001)				
	Unpublished (2001)				
	Contact: Brendan J Loftus				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0208				
	Fax: 301 838 3543				
	Email: b.loftus@tigr.org				
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared				
	DNA library				
	Seq primer: M13-Forward				
	Class: Shotgun				
	High quality sequence start: 17				
	High quality sequence stop: 880.				
FEATURES	Location/Qualifiers				
Source	1..890				
	/organism="Entamoeba histolytica"				
	/strain="HMI:IMSS"				
	/db_xref="taxon:5759"				
	/clone_lib="Entamoeba histolytica Sheared DNA"				
	/note="vector: pHO31; Site_1: Bst I; Constructed at The				
	Institute for Genomic Research (TIGR), Rockville, MD.				
	Genomic DNA isolated from broth cultures of E. histolytica				
	using a method described by Clark and Diamond (Clark, a				
	C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a				
	method for isolate identification. Exp. Parasitol.				
	77:450.). The DNA was mechanically sheared to give a				
	tight size distribution (~2 kb). The v + i method used for				
	the library construction is described in detail in Smith,				
	H.O. and Venter, J.C. (Making small insert libraries for				
	whole genome shotgun sequencing projects. In Genome				
	Sequencing: A Practical Approach, eds. M. Vaudin and B.				
	Barell, Oxford University Press, 1999)."				
BASE COUNT	404 a 65 c 285 g 136 t				
ORIGIN					
Query Match	7.6%; Score 104.4; DB 17; Length 890;				
Best Local Similarity	45.6%; Pred. No. 1.8e-12;				
Matches 369; Conservative	0; Mismatches 441; Indels 0; Gaps 0;				
QY 223	TCTCGGTGAGGACAGAGGTAAGTCATCGTCACCCAGACAGACATCGTCTACAC 282				
Db 841	TTCTTCGTGCTCTAAATCAATTCATCATCGTCTTCTTAATCAATTCATCATCGTC 782				
QY 283	TCCATGTACGAGTCTCTTCATCTCTCTTGGATCAGATCAACAAGGTGGTCTCCAC 342				
Db 781	TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 722				
QY 343	TTGCCAGGTTACCATCATCGATCCGTCGTCAGAACATCCGGTTGGTCCATCGGTATC 402				
Db 721	TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 662				
QY 403	ATCTCCAACTTCTCGTCTTCCACCTGAACAGACAGGAGCTCCGAGCAGTCCATCAAC 462				
Db 661	TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 602				
QY 463	TTTCTCCTAGCAGCATCTCCAAACAGCTCTCGTTTACAACAAGTGGTTCTTCGTACCGTC 522				

```
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHO51; Site:1; Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      434 a   16 c   289 g   169 t
ORIGIN
Query Match      7.6%; Score 104.4; DB 17; Length 908;
Best Local Similarity 45.8%; Pred. No. 1.8e-12;
Matches 360; Conservative 0; Mismatches 426; Indels 0; Gaps 0;
QY      4  TTCACGATGACCAATCCCAATTCACATCTCTCTCTACCAACAACATCCCTGTTGAAGGAC 63
Db      19  TTCATCATCTTCATCATCTTCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 760
QY      64  ATCATCAACGAGTACTTCACACATCAACAGGACTCCCAAGTCTCTGCTCCGAGACCGT 123
Db      759  TTCTTCATCATCTTCATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTC 700
QY      124  AAGAACACCTTGTCGACACCTCCGGTTACACCGCGAGGTCTCCGAGGAGGTGACGTC 183
Db      699  ATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 640
QY      184  CAGCTGAACCAATCTCCCATTCGACTCAAGCTGGGTCTCTCGGTGAGGACAGGT 243
Db      639  GTCGTCTTCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTC 580
QY      244  AAGTCATCTGTCACCCAGACGAAATCGTCTACAACTCCCATGTCAGGACGCTCTCTCC 303
Db      579  ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCATC 520
QY      304  ATCTCTCTCTGATCAGAAATCAAGTGGGTCTCCAACTTCCAGGTACACCATCATC 363
Db      519  ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 460
QY      364  GACTCTGTCAGACAACTCCGGTTGTCATCGGTATCATCTCCAACTCTCTGTCCTTC 423
Db      459  TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 400
QY      424  ACCCTGAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTCCAGCAATCCCAAC 483
Db      399  TTCATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 340
QY      484  AACGCTCTGTTTAAACAAGTGGTCTTCTGTCACCGTCCACCAACATCATGATGGTAAC 543
Db      339  ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 280
QY      544  ATGAAGATCTACATCAACGATAGCTGATCGACCATCATCAAGTCAAGGATGACCGGT 603
Db      279  TTCATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 220
QY      604  ATCAACTCTCCAGACATCACTCTGAGATCAACAAGATCCAGACCGGTCTCTGATC 663
Db      219  ATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 160
QY      664  ACCTCCGACTCCGACATCAACATGATGGATCCGTCGATCTCTCTCTCTCTCTCTCTCTCTCTCT 723
Db      159  GTCCTCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 100
QY      724  TTGACGGTGAAGACATCAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 783
Db      99  TTCATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 40
```

```
QY      784  GACTAC 789
Db      39  ATCTTC 34

RESULT 6
AZ550256
LOCUS    ENT5E58FR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
DEFINITION genomic DNA sequence.
ACCESSION AZ550256
VERSION    AZ550256.1 GI:11175557
KEYWORDS   GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1 (bases 1 to 905)
AUTHORS     Loftus B., Van Aken S. and Fraser C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
JOURNAL     HM1:IMSS sheared DNA library
COMMENT     Unpublished (2000)
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjoftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 100
            High quality sequence stop: 872.
            Location/Qualifiers
                location=1..905
                /organism="Entamoeba histolytica"
                /strain="HM1:IMSS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /notes="Vector: pHO51; Site:1; Bst I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                tight size distribution (~2 kb). The v + i method used for
                the library construction is described in detail in Smith,
                H.O. and Venter, J.C. (Making small insert libraries for
                whole genome shotgun sequencing projects. In Genome
                Sequencing: A Practical Approach, eds. M. Vaudin and B.
                Barell, Oxford University Press, 1999)."
BASE COUNT    203 a   260 c   45 g   396 t
ORIGIN
Query Match      7.6%; Score 104.2; DB 17; Length 905;
Best Local Similarity 48.2%; Pred. No. 2e-12;
Matches 326; Conservative 0; Mismatches 348; Indels 3; Gaps 1;
QY      270  CATCGTCTACAACATCCATGATGAGTCCCTTCTCCATCTCTCTGATCAGATCAACAA 329
Db      150  CGTCATCTTCATCAACATCTTCGTCATCTTCATCATCTTCATCATCTTCATCATCAT 209
QY      330  GTGGGTCTCAACTGCCAGGTACACCATCATCGACTCGGTCAAGCAAACTCCGGTGG 389
Db      210  CTTCTCTCTTCATCTTCATCATCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
QY      390  GTCCATCGGTATCATCTCCAACTTCCTGGTCTTCACCTCGAAGCAAGAGGACTCCGA 449
Db      270  CTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTCT 329
```

FEATURES	source
align quality sequence scop: 740.	
Location/Qualifiers	
1. 976	
/organism="Entamoeba histolytica"	
/strain="HMI:INSS"	
/db_xref="taxon:5759"	
/clone_lib="Entamoeba histolytica"	
/notes="Vector: phos1; site 1; Bst"	

RESULT 8	AZ551618	843 bp	DNA	linear	GSS 14-NOV-2000
LOCUS	AZ551618				
DEFINITION	ENTOV54TR Entamoeba histolytica sheared DNA				
	genomic, DNA sequence.				
ACCESSION	AZ551618				
VERSION	AZ551618.1				
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				



[illegible]

RESULT 10	
AZ546009	
LOCUS	849 bp DNA linear GSS 14-NOV-2000
DEFINITION	Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION	AZ546009
VERSION	AZ546009.1 GI:11167130
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica.
ORGANISM	Entamoeba histolytica
REFERENCE	Eukaryota; Entameobidaae; Entamoeba. 1 (bases 1 to 849)

AUTHORS	Loftus, B., Van Aker, S. and Fraser, C.
TITLE	Determination of clone end sequences from Entamoeba histolytica
JOURNAL	HMI:INSS sheared DNA library
COMMENT	Unpublished (2000) Contact: Brendan J. Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b1ofusts@tigr.org Clones are derived from the Entamoeba histolytica HMI:INSS sheared DNA library Seq primer: M13-Forward Class: shotgun High quality sequence start: 26 High quality sequence stop: 796. Location/Qualifiers 1. .849
FEATURES	source

FEATURES	SOURCE
1. <b>Feature 1</b>	Source 1
2. <b>Feature 2</b>	Source 2
3. <b>Feature 3</b>	Source 3
4. <b>Feature 4</b>	Source 4
5. <b>Feature 5</b>	Source 5
6. <b>Feature 6</b>	Source 6
7. <b>Feature 7</b>	Source 7
8. <b>Feature 8</b>	Source 8
9. <b>Feature 9</b>	Source 9
10. <b>Feature 10</b>	Source 10
11. <b>Feature 11</b>	Source 11
12. <b>Feature 12</b>	Source 12
13. <b>Feature 13</b>	Source 13
14. <b>Feature 14</b>	Source 14
15. <b>Feature 15</b>	Source 15
16. <b>Feature 16</b>	Source 16
17. <b>Feature 17</b>	Source 17
18. <b>Feature 18</b>	Source 18
19. <b>Feature 19</b>	Source 19
20. <b>Feature 20</b>	Source 20
21. <b>Feature 21</b>	Source 21
22. <b>Feature 22</b>	Source 22
23. <b>Feature 23</b>	Source 23
24. <b>Feature 24</b>	Source 24
25. <b>Feature 25</b>	Source 25
26. <b>Feature 26</b>	Source 26
27. <b>Feature 27</b>	Source 27
28. <b>Feature 28</b>	Source 28
29. <b>Feature 29</b>	Source 29
30. <b>Feature 30</b>	Source 30
31. <b>Feature 31</b>	Source 31
32. <b>Feature 32</b>	Source 32
33. <b>Feature 33</b>	Source 33
34. <b>Feature 34</b>	Source 34
35. <b>Feature 35</b>	Source 35
36. <b>Feature 36</b>	Source 36
37. <b>Feature 37</b>	Source 37
38. <b>Feature 38</b>	Source 38
39. <b>Feature 39</b>	Source 39
40. <b>Feature 40</b>	Source 40
41. <b>Feature 41</b>	Source 41
42. <b>Feature 42</b>	Source 42
43. <b>Feature 43</b>	Source 43
44. <b>Feature 44</b>	Source 44
45. <b>Feature 45</b>	Source 45
46. <b>Feature 46</b>	Source 46
47. <b>Feature 47</b>	Source 47
48. <b>Feature 48</b>	Source 48
49. <b>Feature 49</b>	Source 49
50. <b>Feature 50</b>	Source 50
51. <b>Feature 51</b>	Source 51
52. <b>Feature 52</b>	Source 52
53. <b>Feature 53</b>	Source 53
54. <b>Feature 54</b>	Source 54
55. <b>Feature 55</b>	Source 55
56. <b>Feature 56</b>	Source 56
57. <b>Feature 57</b>	Source 57
58. <b>Feature 58</b>	Source 58
59. <b>Feature 59</b>	Source 59
60. <b>Feature 60</b>	Source 60
61. <b>Feature 61</b>	Source 61
62. <b>Feature 62</b>	Source 62
63. <b>Feature 63</b>	Source 63
64. <b>Feature 64</b>	Source 64
65. <b>Feature 65</b>	Source 65
66. <b>Feature 66</b>	Source 66
67. <b>Feature 67</b>	Source 67
68. <b>Feature 68</b>	Source 68
69. <b>Feature 69</b>	Source 69
70. <b>Feature 70</b>	Source 70
71. <b>Feature 71</b>	Source 71
72. <b>Feature 72</b>	Source 72
73. <b>Feature 73</b>	Source 73
74. <b>Feature 74</b>	Source 74
75. <b>Feature 75</b>	Source 75
76. <b>Feature 76</b>	Source 76
77. <b>Feature 77</b>	Source 77
78. <b>Feature 78</b>	Source 78
79. <b>Feature 79</b>	Source 79
80. <b>Feature 80</b>	Source 80
81. <b>Feature 81</b>	Source 81
82. <b>Feature 82</b>	Source 82
83. <b>Feature 83</b>	Source 83
84. <b>Feature 84</b>	Source 84
85. <b>Feature 85</b>	Source 85
86. <b>Feature 86</b>	Source 86
87. <b>Feature 87</b>	Source 87
88. <b>Feature 88</b>	Source 88
89. <b>Feature 89</b>	Source 89
90. <b>Feature 90</b>	Source 90
91. <b>Feature 91</b>	Source 91
92. <b>Feature 92</b>	Source 92
93. <b>Feature 93</b>	Source 93
94. <b>Feature 94</b>	Source 94
95. <b>Feature 95</b>	Source 95
96. <b>Feature 96</b>	Source 96
97. <b>Feature 97</b>	Source 97
98. <b>Feature 98</b>	Source 98
99. <b>Feature 99</b>	Source 99
100. <b>Feature 100</b>	Source 100

1. Dec 849  
/organism="Entamoeba histolytica"  
/strain="HMI:IMMS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="vector: PHO1; site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaubin and B.  
Barell. Oxford University Press, 1999)."

BASE COUNT	199 a	236 c	43 g	371 t
ORIGIN				

Query Match 7.3%; Score 99.4; DB 17; Length 849;  
Best Local Similarity 47.7%; Pred. No. 2.2e-11;  
Matches 326; Conservative 0; Mismatches 351; Indels 6; Gaps 1;

QY	270	CATCGCTACAACCTCCATGATACGAGTGGTTCGATCGATCGATGAGATCAAGAA	329
Db	49	CGTCATCTTCATCAACATCTTCGTCATCTTCATCATCTTCATCATCTTCATCAT	108
QY	330	GTGGGTCTCCAACTTGCAGGTTACACCATCATCGATCGTCAAGAACAACTCCGGTTC	389
Db	109	CTTCTCTTTCATCTTTCATCATCATCTTCTTCTTCTTCTTCATCATCTTCATCTCT	168
QY	390	GTCCATCGGTATCATCTCCAACTCCCTGGTCTTCACCTGAAGCAGAAGGAGATCCGA	449
Db	169	CTTCATCTTCACTTCATCTTTCATCTTCTAAATTCAAATTCATCATCTTCGTCTCTT	228
QY	450	GCAGTCCATCAAACTTCTCCTACGACATCTCCAAAGCTCGGTACAAACAAGTGTT	509
Db	229	CGTCTTCATCATCATCTTTCATCTTCTTCATCTTCTAAATTCAAATTCATCATCTTCT	288
QY	510	CTTCGTCAACGGTCACACACACATGA-----TGGGTAAACATGAAGATCTACATCAACGG	563
Db	289	CTTCTTCATCTTCATCATCATCATCATCATCTTCATCATCTTCATCATCTTCATCATCTT	348
QY	564	TAAGTGTATCGACACACCAACCAAGGTCAGCGGTATCAACTTCTCCAAAGACCAT	623
Db	349	CATCATCTTTCATCATCTTTCATCATCTTCATCATCATCATCATCTTCTTCATCTTCTT	408
QY	624	CACCTTCGAGATCAACAGATCCACAGACACCGGTTCATCACTCCGAGATCCGACAACAT	683
Db	409	CATCTTCTTCATCTTCTTAATTCAAATTCATCATCATCATCTTCTTCTTCGTCTTCTCAT	468
QY	684	CAACATGTGGATCCGTGACTTCTACATCTTCGCCAAGGAGTTGGACGGTAAGACATCAA	743









Seq primer: M13-Forward

Class: Shotgun

High quality sequence start: 20

High quality sequence stop: 890.

#### FEATURES

source  
1..900  
/organism="Entamoeba histolytica"  
/clone="HMI:IMSS"  
/strain="HMI:IMSS"

/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."

BASE COUNT 376 a 63 c 235 g 226 t

#### ORIGIN

Query Match 7.1%; Score 97.4; DB 17; Length 900;  
Best Local Similarity 49.0%; Pred. No. 6.2e-11;  
Matches 260; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 251 TGTGACCCAGAGAGACATCGTTACAACTCCATGTACGAGTCCCTTCCTCCATCTCT 310

DB 860 TCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCG 801

QY 311 TCTGATCAGATCAACAAGTGGTCTCCACATGCGCAGGTACACCATCATCGATCCG 370

DB 800 TCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCG 741

QY 371 TCAAGAACAACTCCGGTGTGTCATCGTATCATCTCCAACTTCCTGGTCTTCACCTCGA 430

DB 740 TCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCG 681

QY 431 AGCAAGAGAGACTCCGAGAGTCCATCACTCTCTAGGACATCTCCAAAGGCCTC 490

DB 680 TCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCG 621

QY 491 CTGGTTACAAAGTGGTCTTCGTCACGTCACCAACATGATGGTAAACATGAAGA 550

DB 620 TCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCT 561

QY 551 TCTAGATCAAGGTAAGTGTATGACAGACATCAAGGTCAAGGTGACCGGTATCAACT 610

DB 560 TCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCA 501

QY 611 TCTCCAGACATCACTCCGTCAGATCAACAAGATCCGACACACGCGTCTCATCCTCCG 670

DB 500 TCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCTTCA 441

QY 671 ACTCGCAACATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 730

DB 440 TCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCAT 381

QY 731 GTAAGGACATCAACATCTGTTCACTCTTCGAGTACACCAAGCTGCTCA 781

DB 380 TCATCGTCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCA 330

#### RESULT 15

BJ324597

LOCUS

DEFINITION BJ324597 Dictyostelium discoideum cDNA library, AF Dictyostelium

discoideum cDNA clone dda7g04 5', mRNA sequence.

ACCESSION BJ324597

BJ324597.1 GI:19154727  
EST.  
Dictyostelium discoideum.  
Dictyostelium discoideum  
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
1 (bases 1 to 797)  
Drushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.  
Full length cDNA of Dictyostelium discoideum at the aggregation stage  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshinigenes.nig.ac.jp.  
Location/Qualifiers

#### JOURNAL

##### COMMENT

/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="dda7g04"  
/clone\_lib="Dictyostelium discoideum cDNA library, AF"  
/sex="mat A"  
/dev\_stage="Aggregation stage"

#### FEATURES

source

1..797

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

/dev\_stage="Aggregation stage"

QY	1019	CCTACAACCTGTTTCATGAAGAACGAGACACATGTACGCCGACACACCTCTCACCGAGGACA	1078
Db	632	TGCGCCACCTCTCAATTTTCATTACAATGCACCAACACAAATCCAAACCAACAACAACA	691
QY	1079	TCTACGCCCATCGGCTCGGTGAGCAGACCAAGGACATCAACGACAACA	1126
Db	692	ACAACAACCAACAAAATTCAAATTCACAACAACAACCCACCAAAAATCCA	739

Search completed: November 7, 2002, 13:45:20  
Job time : 1959 secs